Evolution and regulation of the Z chromosome

by

Vincent Kiplangat Bett

September, 2025

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:
Carl-Philipp Heisenberg, Defense Chair
Beatriz Vicoso, Supervisor
Daniel Zilberman, Internal Committee member
Qi Zhou, External Committee member

The thesis of Vincent Kiplangat Bett, titled *Evolution and regulation of the Z chromosome*, is approved by:

Supervisor: Beatriz Vicoso, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Daniel Zilberman, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Qi Zhou, Center for Evolutionary & Organismal Biology, Zhejiang University, China

Signature: _____

Defense Chair: Carl-Philipp Heisenberg, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I accept full responsibility for the content and factual accuracy of this work, including the data and their analysis and presentation, and the text and citation of other work.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Vincent Kiplangat Bett
September 2025

Signed page is on file

Abstract

Males and females of many species differ in morphology, physiology, and behavior. In taxa with genetic sex determination, sexual differentiation arises largely from sex-biased gene expression, which varies across tissues, developmental stages, and lineages. Increasing evidence highlights chromatin configuration, which can exist in open or closed states, and can be shaped by sex-determination pathways, as a key regulatory layer of this dimorphism. Degeneration of the Y or W chromosome further contributes to sex-specific differences by altering gene copy numbers relative to autosomes in heterogametic sex. To mitigate these imbalances, many eukaryotes have independently evolved dosage compensation mechanisms, often mediated through chromatin-level regulation. In this thesis, we investigate the evolutionary dynamics of sex chromosome differentiation in two species, *Artemia franciscana* and *Cameraria ohridella*, with a particular focus on the extent of dosage compensation following gene loss in the heterogametic sex and the potential chromatin-based mechanisms underlying this process. We further characterize sex-biased gene expression and its regulation through histone modifications. Our analyses also reveal that the *A. franciscana* genome is highly repetitive, with many genes containing intronic transposable elements. We find that enrichment of histone modifications associated with constitutive heterochromatin, positively correlates with variation in gene expression levels. Collectively, these findings underscore role of chromatin regulation in shaping the evolution of sex chromosomes and sexual differentiation.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Professor Beatriz Vicoso, for her invaluable guidance, patience, and continuous support throughout my PhD journey. Her scientific enthusiasm, and meticulous attention to detail have been fundamental in shaping this dissertation. I am also indebted to the members of my thesis committee, Professor Daniel Zilberman and Professor Qi Zhou, for their valuable insights during the course of this research. Furthermore, I wish to acknowledge the members of the Vicoso group for their collaboration, fruitful discussions, and contributions, which have considerably enriched this work. I am also grateful to the ISTA Scientific Computing Unit for their resources and technical support.

I extend my sincere thanks to my family and friends for their steadfast encouragement and support during this PhD journey.

This work was supported by the Austrian Science Fund (FWF) through grants PAT8748323 and SFB F88-10 awarded to Professor Beatriz Vicoso.

About the Author

Vincent Kiplangat Bett obtained his BSc. in Biomedical Science and Technology from Egerton University, Kenya, in 2015. He then joined the Kenya Medical Research Institute, serving first as an intern and later as a research assistant. In 2017, he enrolled at the University of Cambridge (UK) for an MPhil in Biological Science (Wellcome Sanger Institute), where he investigated the epidemiology and transmission of *Staphylococcus aureus* using One Health approach. In 2019, Vincent joined the graduate school as a PhD student and affiliated with the Beatriz Vicoso in 2020 to work on the evolutionary dynamics of Z chromosome and its regulation. During his PhD studies, he presented his research at the 3rd Joint Congress on Evolutionary Biology in Montreal, Canada (2024), and the Congress of the European Society for Evolutionary Biology (ESEB) in Barcelona, Spain (2025). Vincent has published some of his research findings in Genome Biology and Evolution (GBE) and Molecular Biology and Evolution (MBE).

List of Collaborators and Publications

The following two publications form the basis of this thesis and which have been fully re-used in Chapters 2 and 3.

Bett V.K., Macon A., Vicoso B.*, Elkrewi M.* 2024. Chromosome-Level Assembly of *Artemia franciscana* Sheds Light on Sex Chromosome Differentiation. *Genome Biology and Evolution*, 2024, Jan, 16(1): evae006

Bett V.K.*, Trejo-Arellano M.S, Vicoso B.* 2025. Chromatin Landscape Is Associated With Sex-Biased Expression and Drosophila-Like Dosage Compensation of the Z Chromosome in *Artemia franciscana*, *Molecular Biology and Evolution*, 2025, May, 42(5):msaf085

Publications in preparation

Bett V.K.*, Macon A., Elkrewi M., Vicoso B.* 2026. Unravelling Sex Chromosome Evolution and Dosage Compensation Mechanisms in *Cameraria ohridella* (horse-chestnut leaf miner moth)

Publications not included in the thesis

Elkrewi M, Khauratovich U., Toups M.A, **Bett V.K**, Mrnjavac A., Macon A., Fraisse C., Sax L, Huylmans A.K, Hontoria F., Vicoso B. ZW sex-chromosome evolution and contagious parthenogenesis in Artemia brine shrimp, Genetics, 2022, October, 222(2):iyac123

Table of Contents

Contents

Chapter 1 : Introduction

Evolutionary Dynamics of Sex Chromosome Evolution

Sex chromosomes are derived from pairs of initially homologous autosomes and have evolved independently multiple times across the eukaryotic tree of life (Bachtrog et al. 2011) and there are two types: XY (male heterogamety) and ZW (female heterogamety). The process typically begins when an autosome acquires a sex-determining gene (D. Charlesworth et al. 2005), after which sexually antagonistic mutations accumulate around this locus. Chromosomal inversions can then arise on the proto-Y (or W) chromosome, preventing the inverted segment from aligning with its homologous region on the proto-X during meiosis. Each successive inversion further expands the non-recombining portion of the sex chromosomes. These regions, known as strata, differ in both their age and the degree of genetic degeneration, depending on when the inversions took place (D. Charlesworth et al. 2005; Lahn and Page 1999). Over time, the non-recombining regions on the newly formed Y (or W) chromosome accumulate mutations, deletions, and insertions of transposable elements, which ultimately lead to gene loss (Bachtrog 2013). In certain lineages, sex chromosomes are subject to chromosomal rearrangements such as fusions, fissions, translocations, and turnovers, contributing to the evolutionary variation observed in sex chromosome systems (Toups and Vicoso 2025). One notable example is the formation of neo-sex chromosomes, which may result from fusions between sex chromosomes and autosomes or through the recruitment of B chromosomes, where dispensable elements are co-opted into sex-specific roles. These neo-sex chromosomes have independently evolved across various taxa, including multiple species within the *Drosophila* genus (Bachtrog 2013). The timing of recombination suppression between the segregating chromosomes gives rise to distinct evolutionary strata, offering valuable insights into the dynamics of sex chromosome evolution. For instance, in *D. albomicans*, neo-Y gene expression declines within 0.1 million years. After one million years, the neo-Y in *D. miranda* exhibits substantial gene decay, with nearly 40% of its genes disrupted by frameshift mutations or premature stop codons, while in *D. pseudoobscura*, the neo-Y has lost most of its genes and become largely heterochromatic after 15 million years, resulting in an architecture that closely resembles that of the ancestral Y chromosome in *Drosophila* (Vicoso 2019; Bachtrog 2013).

Neo-sex chromosomes are relatively common in Lepidoptera, with the most of the taxa exhibiting the putative ancestral haploid karyotype (n = 31) (de Vos et al. 2020). These chromosomal rearrangements occur more frequently than would be expected based on a null model (Anderson et al. 2020), with at least 30 independent Z fusions documented (Wright et al. 2024). W chromosomes have arisen independently on multiple occasions across various lineages (Blackmon et al. 2017), underscoring their dynamic role in sex chromosome evolution. The occurrence and functional significance of these chromosomes are not consistent across Lepidoptera lineages. For instance, no definitive involvement of a W or neo-W chromosome in sex determination or reproduction has been identified in *Samia cynthia* subspecies (Yoshido et al. 2016). Conversely, the W chromosome in the silkworm *Bombyx mori* contains the dominant feminizing gene, *Fem*, which is crucial for promoting female development (Kiuchi et al. 2014). These lineage-specific differences indicate that W and neo-W chromosomes may undergo distinct evolutionary pathways; however, the broader molecular and regulatory mechanisms driving their evolution remain largely unexplored.

Molecular Basis of Sexual Dimorphism

Males and females across numerous species often exhibit striking differences in morphology, physiology, and behavior despite sharing an almost identical genome (Parsch and Ellegren

2013). In species with environmental sex determination (ESD), genetic differences between the sexes are completely absent. Even in systems with genetic sex determination, the differences may be limited to a few genes on a sex-specific chromosome such as the Y chromosome in male heterogamety or the W chromosome in female heterogamety. How do such profound phenotypic differences arise between the sexes when their underlying genetic material is largely similar? Genome-wide transcriptomic studies across a broad range of taxa including nematodes, insects, birds, and mammals, demonstrate a strong correlation between sex-biased gene expression and patterns of sexual dimorphism, suggesting that gene expression profiles may serve as molecular proxies for phenotypic differences between the sexes. In birds, the tissues exhibiting greater degrees of sexual dimorphism also display higher levels of sex-biased gene expression (Mank et al. 2008). Similarly, previous studies in Drosophila and birds have shown that differences in RNA abundance between sexes become more pronounced with successive developmental stages (Perry et al. 2014; Mank et al. 2010).

The molecular mechanisms underlying sexual dimorphism are remarkably diverse across the animal kingdom, ranging from splice-based, cell-autonomous sex determination systems in insects to gonad-dependent hormonal signaling pathways in mammals (McKeown 1992). Despite this mechanistic diversity, many of these pathways converge on the transcriptional activity of the *doublesex/mab-3 related* (*Dmrt*) family of genes, which play a central role in initiating sex-specific development across a wide array of animal taxa (Kopp 2012). *Dmrt* genes function as key developmental regulators, integrating spatial, temporal, and sex-specific cues to guide cells toward male or female fates. In insects, *Dmrt* genes operate through an alternative splicing cascade. The gene, doublesex (*dsx*), for example, produces sex-specific isoforms that drive the differentiation of sexually dimorphic traits: the male-specific isoform promotes male trait development while repressing female characteristics, and vice versa for the female-specific isoform (Kopp 2012). In some vertebrates, *Dmrt* genes may even act as the primary sex-determining factors, highlighting their evolutionary significance and functional versatility in governing sexual differentiation across metazoans. Epigenetic modifications can also modulate *Dmrt* gene activity during sex determination. For instance, in turtles with temperature-dependent sex determination, DNA methylation at the *dmrt1* promoter influences its expression and can induce reversible shifts in sexual phenotype (Ge et al. 2018). However, not all sex-biased genes are direct targets of canonical sex-determination pathways. Instead, they may be regulated by additional mechanisms involving chromatin accessibility, non-coding RNAs, DNA methylation, or histone modifications. In *Daphnia pulex*, a species in which males and females share identical genomes, sex-specific epigenetic landscapes correlate with differential expression of genes tied to both physiological traits (e.g., metabolism and lifespan) and morphological features (e.g., antennae and body size) (Kvist et al. 2020). Recent work on two closely related *Drosophila* species (*Drosophila melanogaster* and *Drosophila simulans*) has shed light on the potential role of chromatin state in regulating sex-biased gene expression, particularly through influence of histone modifications (Nanni et al. 2023). This study investigated three histone modifications; H3K4me3, associated with transcriptionally active or "open" chromatin and H3K27me2/H3K27me3, linked to repressive or "closed" chromatin. The findings support the "Open in Same Sex and/or Closed in Opposite" model (Nanni et al. 2023), a framework originally introduced by Brown and Bachtrog (2014) (Brown and Bachtrog 2014). This model suggests that sex-biased gene expression is influenced by sex-specific chromatin accessibility, wherein genes remain epigenetically open in the sex where they are actively transcribed (Nanni et al. 2023). Notably, the scope of this study was limited to three histone marks and focused solely on head tissue. A more extensive epigenomic analysis encompassing diverse tissues, a broader array of histone modifications, and a wider variety of species is

required to comprehensively evaluate sex-specific chromatin dynamics underlying molecular sexual dimorphism.

Epigenetic Regulation: Histone Modifications and Chromatin Structure

The term *epigenetics* was first coined in 1942 by Conrad Waddington to describe factors affecting the interaction between genes and their products that result in phenotypic changes (Waddington 2012). With the development of new molecular techniques, the definition has expanded to encompass all chemical modifications of DNA or histone proteins within chromatin that are heritable through cell division. These modifications, which occur without altering the DNA sequence, affect *gene expre*ssion by regulating chromatin structure and DNA accessibility (Deans and Maggert 2015). In model species such as humans, mice, and *Drosophila*, major mechanisms include histone modifications, chromatin remodeling, DNA methylation, and non-coding RNAs (Gibney and Nolan 2010; Nègre et al. 2011; Dunham et al. 2012). In addition to their involvement in transcriptional regulation, epigenetic processes contribute to the maintenance of genomic stability (Feng and Riddle 2020).

A central feature of epigenetic regulation involves the structural organization of DNA within the nucleus. DNA is packaged by histone protein octamers to form a nucleoprotein complex known as chromatin. The fundamental subunits of chromatin, known as nucleosomes, comprise 147 base pairs of DNA coiled around a core histone octamer that includes paired H2A, H2B, H3, and H4 proteins. Nucleosomes are linked via stretches of linker DNA, forming a continuous polymer stabilized by the H1 histone protein (Quina et al. 2006). Critical processes of the cells, including replication, transcriptional regulation, DNA repair, recombination, heterochromatin assembly and chromosome condensation, rely on how the chromatin structure is organized (Kouzarides 2007). This organization is modulated by post-translational modifications of the N-terminal tails of histone proteins, including methylation, acetylation, ubiquitination, sumoylation, phosphorylation, and poly(ADP-ribosylation) (Kouzarides 2007). There are more than 60 different post-translational modifications that have been identified in the histone tails (Chan and Maze 2020). These covalent modifications can change histone charge, affecting chromatin structure and DNA binding, or provide sites for chromatin-associated proteins (Quina et al. 2006). These modifications may function in an antagonistic or synergistic manner. Conserved lysine residues on the N-termini of H3 and H4 histones are primary modification targets. Histone acetylation, which neutralizes the positive charge of lysines, weakens histone-DNA interactions and promotes an open chromatin state (euchromatin), allowing access to nucleases and transcription factors for gene expression (Kouzarides 2007). In contrast, histone methylation has context-dependent effects: methylation at certain residues (e.g., H3K4me3) is associated with transcriptional activation, whereas methylation at other sites (e.g., H3K9me3, H3K27me3) promotes chromatin compaction (heterochromatin) and gene silencing (Arney and Fisher 2004). Heterochromatin is classified as facultative, marked by mono- and di-methylation of H3/H4 tails in transcriptionally inactive regions, or constitutive, found in gene-poor, repeat-rich regions such as pericentromeres, centromeres, and sex chromosomes (Arney and Fisher 2004; Quina et al. 2006).

Epigenetics of Dosage Compensation

As genes are lost from degenerating (neo-)Y or W chromosomes, gene expression imbalances occur in XY males or ZW females, who are left with only a single functional copy of X- or Z-linked genes, while still retaining two copies of autosomal genes. This imbalance is thought to be deleterious (Veitia 2005), prompting evolution of molecular mechanisms targeting the X or Z chromosomes that restore their ancestral expression levels. The primary function of these mechanisms is to restore equilibrium between sex chromosomes and autosomes (a process

termed dosage compensation) as such imbalances disrupt gene networks and reduce organism's fitness, making them targets of strong natural selection. As a secondary consequence, they also equalize X- or Z-linked gene expression between males and females (dosage balance), which likely evolves indirectly. In some species, dosage compensation evolved to affect the entire chromosome. However, partial dosage compensation exists in some clades, where gene-by-gene dosage balancing between the sexes occurs for dosage-sensitive genes, but other X/Z-specific genes remain expressed at different levels (Gu and Walters 2017). Although ZW sex-determination systems are common among animals, their molecular and evolutionary dynamics are not well understood. Most insights about dosage compensation comes from XY system studies, and early ZW research has typically reported only partial, gene-specific compensation. This led to the belief that global dosage compensation is rare or restricted to a highly dosage-sensitive genes in ZW systems (Zimmer et al. 2016). However, recent studies across diverse taxa including monarch butterflies (Gu et al. 2019), free larval cercariae of blood flukes *Schistosoma mansoni* (Picard et al. 2019), and softshell turtles *Apalone spinifera* (Bista et al. 2021) show global dosage compensation, challenging previous assumptions and suggesting greater evolutionary diversity in ZW systems. Still, important subtleties remain: in *S. mansoni*, there are consistent male-biased expression of Z-linked genes, a pattern once interpreted as evidence against a global mechanism; and in *A. spinifera*, dosage compensation appears to operate only under certain environmental conditions.

Three primary mechanisms of chromosome-wide dosage compensation have been described: (i) inactivation of the entire X/Z chromosome in the homogametic sex, as observed in mammals; (ii) reduced transcription from both X/Z chromosomes by half in the homogametic sex, exemplified by *Caenorhabditis elegans* (Lau and Csankovszki 2015); and (iii) upregulation of gene expression in the heterogametic sex, as seen in *Drosophila*. Mechanisms involving downregulation of the X/Z chromosome in the homogametic sex remain intriguing. Ohno proposed a two-step evolutionary process for such mechanisms: initially, the upregulation of X (or Z) expression occurs in a non-sex-specific manner to restore ancestral expression levels in the heterogametic sex; subsequently, downregulation of X/Z in the homogametic sex compensates for any excess expression, thereby equalizing gene expression between males and females (B. Charlesworth 1996).

Dosage compensation is associated with chromatin structural changes, which may involve DNA methylation, non-coding RNAs, and post-translational histone modifications. In mammals, random inactivation of the X-chromosome in females is regulated by the gene *Xist* (*X* inactive-specific transcript). *Xist* produces a long non-coding RNA that initiates X silencing by recruiting various complexes and factors that modify the chromatin structure (Brockdorff et al. 2020). The resulting epigenetic changes include DNA methylation at CpG promoters, reductions in transcription-associated histone marks (H3K9ac, H3K4me1), and increased enrichment of repressive marks (H3K27me3, H2AK119Ub) on the inactivated X chromosome in females (Muyle et al. 2021). In *Drosophila melanogaster,* the male X chromosome undergoes global hyperacetylation, marked by increased H4K16ac histone modification (Conrad and Akhtar 2012). This change is mediated by the Male-Specific Lethal (MSL) complex, composed of at least five core proteins; MSL-1, MSL-2, MSL-3, Maleless (MLE), and Males absent on the First (MOF) along with two noncoding RNAs, roX1 and roX2 (MUKHERJEE and BEERMANN 1965). Functional disruption of MSL complex components results in impaired development or lethality (Belote and Lucchesi 1980). The process of MSL-mediated dosage compensation in *Drosophila* proceeds in two stages. Initially, the complex is recruited to high-affinity entry sites on the X chromosome that are characterized by GA-rich DNA motifs (Alekseyenko et al. 2008). Subsequently, the complex spreads to actively transcribed regions across the X chromosome. This spreading is facilitated by H3K36

trimethylation, a histone modification marking active chromatin (Larschan et al. 2007), and by three-dimensional interactions between entry sites and other active X-linked loci (Schauer et al. 2017). Through this distribution, the MSL complex upregulates gene expression across the male X chromosome by modifying chromatin structure, specifically through acetylation of histone H4 at lysine 16 (H4K16ac), which is associated with chromatin decompaction and increased transcriptional activity (Akhtar and Becker 2000). Similarly, in green anole lizards, male-specific chromatin machinery enriches H4K16ac across X-linked genes for two-fold upregulation (Marin et al. 2017). In *Anopheles gambiae*, however, upregulation of the male X is mediated by the SOA/007 gene (Kalita et al. 2023; Krzywinska et al. 2023), which likely functions as a transcription factor by binding directly to the promoter regions of the dosage compensated X-linked genes (Toups and Vicoso 2025). In monarch butterflies, both the ancestral Z chromosome and the more recently evolved neo-Z chromosome (formed by a fusion event between the ancestral Z and an autosomal chromosome) exhibited two distinct mechanisms of dosage compensation. Specifically, the neo-Z region displays a two-fold increase in transcription in females, correlated with elevated levels of H4K16ac, a pattern analogous to that observed in Drosophila. Conversely, genes located on the ancestral Z are transcriptionally repressed in males (unknown molecular mechanism) and show reduced H4K16ac enrichment, mirroring nematode-like dosage compensation patterns reported in other Lepidopteran species (Gu et al. 2019).

In this thesis, we generated genomic resources for two ZW sex-determining species, *Artemia franciscana* and *Cameraria ohridella*, to investigate the evolution and regulation of sex chromosomes. We examined the dynamics of sex chromosome differentiation and the extent of dosage compensation following gene loss in the heterogametic sex. We further explored chromatin-based mechanisms of dosage compensation, including the role of histone modifications in regulating sex-biased gene expression. Our analyses also revealed that the *A. franciscana* genome is highly repetitive, with many genes containing intronic transposable elements, and that histone modifications associated with constitutive heterochromatin are enriched at these genes, correlating with their expression levels.

References:

Akhtar, Asifa, and Peter B Becker. 2000. "Activation of Transcription through Histone H4 Acetylation by MOF, an Acetyltransferase Essential for Dosage Compensation in Drosophila." *Molecular Cell* 5 (2): 367–75. https://doi.org/10.1016/S1097-2765(00)80431-1.

Alekseyenko, Artyom A., Shouyong Peng, Erica Larschan, et al. 2008. "A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome." *Cell* 134 (4): 599–609. https://doi.org/10.1016/j.cell.2008.06.033.

Anderson, N. W., Hjelmen, C. E., & Blackmon, H. (2020). "The probability of fusions joining sex chromosomes and autosomes." *Biology letters*, *16*(11), 20200648. https://doi.org/10.1098/rsbl.2020.0648

Arney, Katharine L., and Amanda G. Fisher. 2004. "Epigenetic Aspects of Differentiation." *Journal of Cell Science* (England) 117 (Pt 19): 4355–63. https://doi.org/10.1242/jcs.01390.

Bachtrog, Doris. 2013. "Y-Chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration." *Nature Reviews. Genetics* (England) 14 (2): 113–24. https://doi.org/10.1038/nrg3366.

Bachtrog, Doris, Mark Kirkpatrick, Judith E. Mank, et al. 2011. "Are All Sex Chromosomes Created Equal?" *Trends in Genetics : TIG* (England) 27 (9): 350–57. https://doi.org/10.1016/j.tig.2011.05.005.

Belote, John M, and John C Lucchesi. 1980. "MALE-SPECIFIC LETHAL MUTATIONS OF DROSOPHILA MELANOGASTER." *Genetics* 96 (1): 165–86. https://doi.org/10.1093/genetics/96.1.165.

Bista, Basanta, Zhiqiang Wu, Robert Literman, and Nicole Valenzuela. 2021. "Thermosensitive Sex Chromosome Dosage Compensation in ZZ/ZW Softshell Turtles, Apalone Spinifera." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* (England) 376 (1833): 20200101. https://doi.org/10.1098/rstb.2020.0101.

Blackmon, Heath, Laura Ross, and Doris Bachtrog. 2017. "Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects." *The Journal of Heredity* (United States) 108 (1): 78–93. https://doi.org/10.1093/jhered/esw047.

Brockdorff, Neil, Joseph S. Bowness, and Guifeng Wei. 2020. "Progress toward Understanding Chromosome Silencing by Xist RNA." *Genes & Development* (United States) 34 (11–12): 733–44. https://doi.org/10.1101/gad.337196.120.

Brown, Emily J., and Doris Bachtrog. 2014. "The Chromatin Landscape of Drosophila: Comparisons between Species, Sexes, and Chromosomes." *Genome Research* (United States) 24 (7): 1125–37. https://doi.org/10.1101/gr.172155.114.

Chan, Jennifer C., and Ian Maze. 2020. "Nothing Is Yet Set in (Hi)Stone: Novel Post-Translational Modifications Regulating Chromatin Function." *Trends in Biochemical Sciences* (England) 45 (10): 829–44. https://doi.org/10.1016/j.tibs.2020.05.009.

Charlesworth, B. 1996. "The Evolution of Chromosomal Sex Determination and Dosage Compensation." *Current Biology : CB* (England) 6 (2): 149–62. https://doi.org/10.1016/s0960-9822(02)00448-7.

Charlesworth, D., B. Charlesworth, and G. Marais. 2005. "Steps in the Evolution of Heteromorphic Sex Chromosomes." *Heredity* (England) 95 (2): 118–28. https://doi.org/10.1038/sj.hdy.6800697.

Conrad, Thomas, and Asifa Akhtar. 2012. "Dosage Compensation in Drosophila Melanogaster: Epigenetic Fine-Tuning of Chromosome-Wide Transcription." *Nature Reviews Genetics* 13 (2): 123–34. https://doi.org/10.1038/nrg3124.

Deans, Carrie, and Keith A. Maggert. 2015. "What Do You Mean, 'Epigenetic'?" *Genetics* (United States) 199 (4): 887–96. https://doi.org/10.1534/genetics.114.173492.

Dickson, Barry J. 2008. "Wired for Sex: The Neurobiology of Drosophila Mating Decisions." *Science* 322 (5903): 904–9. https://doi.org/10.1126/science.1159276.

Dunham, Ian, Anshul Kundaje, Shelley F. Aldred, et al. 2012. "An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature* 489 (7414): 57–74. https://doi.org/10.1038/nature11247.

Feng, Justina X., and Nicole C. Riddle. 2020. "Epigenetics and Genome Stability." *Mammalian Genome : Official Journal of the International Mammalian Genome Society* (United States) 31 (5–6): 181–95. https://doi.org/10.1007/s00335-020-09836-2.

Ge, Chutian, Jian Ye, Ceri Weber, et al. 2018. "The Histone Demethylase KDM6B Regulates Temperature-Dependent Sex Determination in a Turtle Species." *Science* 360 (6389): 645–48. https://doi.org/10.1126/science.aap8328.

Gibney, E. R., and C. M. Nolan. 2010. "Epigenetics and Gene Expression." *Heredity* (England) 105 (1): 4–13. https://doi.org/10.1038/hdy.2010.54.

Gu, Liuqi, Patrick F. Reilly, James J. Lewis, Robert D. Reed, Peter Andolfatto, and James R. Walters. 2019. "Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly." *Current Biology* 29 (23): 4071-4077.e3. https://doi.org/10.1016/j.cub.2019.09.056.

Gu, Liuqi, and James R Walters. 2017. "Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details." *Genome Biology and Evolution* 9 (9): 2461–76. https://doi.org/10.1093/gbe/evx154.

Kalita, Agata Izabela, Eric Marois, Magdalena Kozielska, et al. 2023. "The Sex-Specific Factor SOA Controls Dosage Compensation in Anopheles Mosquitoes." *Nature* 623 (7985): 175–82. https://doi.org/10.1038/s41586-023-06641-0.

Kiuchi, Takashi, Hikaru Koga, Munetaka Kawamoto, et al. 2014. "A Single Female-Specific piRNA Is the Primary Determiner of Sex in the Silkworm." *Nature* (England) 509 (7502): 633–36. https://doi.org/10.1038/nature13315.

Kopp, Artyom. 2012. "Dmrt Genes in the Development and Evolution of Sexual Dimorphism." *Trends in Genetics : TIG* (England) 28 (4): 175–84. https://doi.org/10.1016/j.tig.2012.02.002.

Kouzarides, Tony. 2007. "Chromatin Modifications and Their Function." *Cell* (United States) 128 (4): 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

Krzywinska, Elzbieta, Paolo Ribeca, Luca Ferretti, Andrew Hammond, and Jaroslaw Krzywinski. 2023. "A Novel Factor Modulating X Chromosome Dosage Compensation in Anopheles." *Current Biology* 33 (21): 4697-4703.e4. https://doi.org/10.1016/j.cub.2023.09.001.

Kvist, Jouni, Camila Gonçalves Athanàsio, Michael E. Pfrender, James B. Brown, John K. Colbourne, and Leda Mirbahai. 2020. "A Comprehensive Epigenomic Analysis of Phenotypically Distinguishable, Genetically Identical Female and Male Daphnia Pulex." *BMC Genomics* 21 (1): 17. https://doi.org/10.1186/s12864-019-6415-5.

Lau, Alyssa C., and Györgyi Csankovszki. 2015. "Balancing up and Downregulation of the C. Elegans X Chromosomes." *Current Opinion in Genetics & Development* (England) 31 (April): 50–56. https://doi.org/10.1016/j.gde.2015.04.001.

Mank, Judith E., Lina Hultin-Rosenberg, Matthew T. Webster, and Hans Ellegren. 2008. "The Unique Genomic Properties of Sex-Biased Genes: Insights from Avian Microarray Data." *BMC Genomics* (England) 9 (March): 148. https://doi.org/10.1186/1471-2164-9-148.

Mank, Judith E., Kiwoong Nam, Björn Brunström, and Hans Ellegren. 2010. "Ontogenetic

Complexity of Sexual Dimorphism and Sex-Specific Selection." *Molecular Biology and Evolution* 27 (7): 1570–78. https://doi.org/10.1093/molbev/msq042.

Marin, Ray, Diego Cortez, Francesco Lamanna, et al. 2017. "Convergent Origination of a Drosophila-like Dosage Compensation Mechanism in a Reptile Lineage." *Genome Research* (United States) 27 (12): 1974–87. https://doi.org/10.1101/gr.223727.117.

McKeown, Michael. 1992. "Sex Differentiation: The Role of Alternative Splicing." *Current Opinion in Genetics & Development* 2 (2): 299–303. https://doi.org/10.1016/S0959-437X(05)80288-6.

MUKHERJEE, ARDHENDU S., and W. BEERMANN. 1965. "Synthesis of Ribonucleic Acid by the X-Chromosomes of Drosophila Melanogaster and the Problem of Dosage Compensation." *Nature* 207 (4998): 785–86. https://doi.org/10.1038/207785a0.

Muyle, Aline, Doris Bachtrog, Gabriel A. B. Marais, and James M. A. Turner. 2021. "Epigenetics Drive the Evolution of Sex Chromosomes in Animals and Plants." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1826): 20200124. https://doi.org/10.1098/rstb.2020.0124.

Nanni, Adalena V, Natalie Martinez, Rita Graze, et al. 2023. "Sex-Biased Expression Is Associated With Chromatin State in Drosophila Melanogaster and Drosophila Simulans." *Molecular Biology and Evolution* 40 (5): msad078. https://doi.org/10.1093/molbev/msad078.

Nègre, Nicolas, Christopher D. Brown, Lijia Ma, et al. 2011. "A Cis-Regulatory Map of the Drosophila Genome." *Nature* (England) 471 (7339): 527–31. https://doi.org/10.1038/nature09990.

Parsch, John, and Hans Ellegren. 2013. "The Evolutionary Causes and Consequences of Sex-Biased Gene Expression." *Nature Reviews Genetics* 14 (2): 83–87. https://doi.org/10.1038/nrg3376.

Perry, Jennifer C., Peter W. Harrison, and Judith E. Mank. 2014. "The Ontogeny and Evolution of Sex-Biased Gene Expression in Drosophila Melanogaster." *Molecular Biology and Evolution* (United States) 31 (5): 1206–19. https://doi.org/10.1093/molbev/msu072.

Picard, Marion A. L., Beatriz Vicoso, David Roquis, et al. 2019. "Dosage Compensation throughout the Schistosoma Mansoni Lifecycle: Specific Chromatin Landscape of the Z Chromosome." *Genome Biology and Evolution* (England) 11 (7): 1909–22. https://doi.org/10.1093/gbe/evz133.

Quina, A. S., M. Buschbeck, and L. Di Croce. 2006. "Chromatin Structure and Epigenetics." *Biochemical Pharmacology* (England) 72 (11): 1563–69. https://doi.org/10.1016/j.bcp.2006.06.016.

Schauer, Tamás, Yad Ghavi-Helm, Tom Sexton, et al. 2017. "Chromosome Topology Guides the Drosophila Dosage Compensation Complex for Target Gene Activation." *EMBO Reports* 18 (10): 1854–68. https://doi.org/10.15252/embr.201744292.

Toups, Melissa A., and Beatriz Vicoso. 2025. "Insect Sex Chromosome Evolution: Conservation, Turnover, and Mechanisms of Dosage Compensation." *Current Opinion in Insect Science*, July 12, 101411. https://doi.org/10.1016/j.cois.2025.101411.

Veitia, Reiner A. 2005. "Gene Dosage Balance: Deletions, Duplications and Dominance." *Trends in Genetics : TIG* (England) 21 (1): 33–35. https://doi.org/10.1016/j.tig.2004.11.002.

Vicoso, Beatriz. 2019. "Molecular and Evolutionary Dynamics of Animal Sex-Chromosome Turnover." *Nature Ecology & Evolution* 3 (12): 1632–41. https://doi.org/10.1038/s41559-019-1050-8.

Vos, Jurriaan M. de, Hannah Augustijnen, Livio Bätscher, and Kay Lucek. 2020. "Speciation through Chromosomal Fusion and Fission in Lepidoptera." *Philosophical Transactions*

of the Royal Society B: Biological Sciences 375 (1806): 20190539. https://doi.org/10.1098/rstb.2019.0539.

Waddington, C. H. 2012. "The Epigenotype. 1942." *International Journal of Epidemiology* (England) 41 (1): 10–13. https://doi.org/10.1093/ije/dyr184.

Wright, Charlotte J., Lewis Stevens, Alexander Mackintosh, Mara Lawniczak, and Mark Blaxter. 2024. "Comparative Genomics Reveals the Dynamics of Chromosome Evolution in Lepidoptera." *Nature Ecology & Evolution* 8 (4): 777–90. https://doi.org/10.1038/s41559-024-02329-4.

Yoshido, A., F. Marec, and K. Sahara. 2016. "The Fate of W Chromosomes in Hybrids between Wild Silkmoths, Samia Cynthia Ssp.: No Role in Sex Determination and Reproduction." *Heredity* 116 (5): 424–33. https://doi.org/10.1038/hdy.2015.110.

Zimmer, Fabian, Peter W. Harrison, Christophe Dessimoz, and Judith E. Mank. 2016. "Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome." *Genome Biology and Evolution* (England) 8 (4): 1233–42. https://doi.org/10.1093/gbe/evw075.

Chapter 2: Chromosome-Level Assembly of *Artemia franciscana* Sheds Light on Sex Chromosome Differentiation

This second chapter is based on work co-authored with my colleague Marwan Elkrewi, and developed under the guidance of our supervisor, Beatriz Vicoso. The study's concept was jointly designed and implemented through our collaboration. My specific contributions are the annotation of repeats and genes in the genome, expression analyses, and participation in the initial stages of genome assembly.

Chapter 2: Chromosome-Level Assembly of *Artemia franciscana* Sheds Light on Sex Chromosome Differentiation

Vincent Kiplangat Bett, Ariana Macon, Beatriz Vicoso[@], and Marwan Elkrewi[@]

Abstract

Since the commercialization of brine shrimp (genus Artemia) in the 1950s, this lineage, and in particular the model species Artemia franciscana, has been the subject of extensive research. However, our understanding of the genetic mechanisms underlying various aspects of their reproductive biology, including sex determination, is still lacking. This is partly due to the scarcity of genomic resources for Artemia species and crustaceans in general. Here, we present a chromosome-level genome assembly of A. franciscana (Kellogg 1906), from the Great Salt Lake, United States. The genome is 1 GB, and the majority of the genome (81%) is scaffolded into 21 linkage groups using a previously published high-density linkage map. We performed coverage and FST analyses using male and female genomic and transcriptomic reads to quantify the extent of differentiation between the Z and W chromosomes. Additionally, we quantified the expression levels in male and female heads and gonads and found further evidence for dosage compensation in this species.

Significance

Besides its economic importance, the unique characteristics of Artemia brine shrimp have made it a great model for exploring many evolutionary questions, including the evolution of sex chromosomes, sexual dimorphism, asexuality, and plasticity of reproductive modes. The genome assembly produced here will be an invaluable resource for advancing the efforts made in elucidating the genetic architecture of evolutionary and biologically relevant traits. It will also pave the way for more comprehensive studies in the phylogenomics and comparative genomics of Arthropods.

Introduction

Artemia brine shrimp are crustaceans belonging to the Anostracan order in the Branchiopoda class, which includes around 1,200 species (Castellucci et al. 2022). They live in saline/hypersaline inland water bodies, with a very wide geographical distribution (Eimanifar et al. 2015). They are very adaptable and can survive in extreme environments; this is facilitated by their ability to produce both live offspring and encapsulated cysts, which can survive in dry conditions for extended periods of time (Criel and Macrae 2002). Their adaptability, ease of rearing, and high nutritional value have made them very popular in the aquarium trade and aquaculture industry (Lavens and Sorgeloos 2000). Artemia has other industrial uses, which include controlling algal growth in salt mines and improving the efficiency of salt production (Van Stappen et al. 2020). Furthermore, they have been extensively used in toxicity and ecotoxicity testing due to their abundance, cost-effectiveness, and ease of manipulation in the laboratory (Nunes et al. 2006; Rajabi et al. 2015). *Artemia franciscana* is arguably the most extensively studied Artemia species; however, to date, the genomic resources for *A. franciscana* are limited to 2 scaffold-level assemblies (De Vos et al. 2021; Jo et al. 2021). While they have yielded important insights into the adaptation to extreme environments, several aspects of their reproductive biology, including the molecular basis of sex determination and the extent of sex chromosome differentiation, are difficult to elucidate without a chromosome-level assembly (Huylmans et al. 2019). Currently, the closest relative with a published chromosome-level assembly is the Asian *Artemia sinica* (Elkrewi et al. 2022), from which *A. franciscana* diverged 30 million years ago (Baxevanis et al. 2006; Maniatsi et

al. 2009). Artemia are also a great model for sex chromosome evolution, as they have ZW sex chromosomes (Bowen 1963; De Vos et al. 2013; Elkrewi et al. 2022; Boyer et al. 2023). Sex chromosomes are known to evolve from autosomes, when one of them acquires a sex determination gene. Recombination is then thought to be lost in a stepwise manner, creating strata of different ages (Lahn and Page 1999; Handley et al. 2004), but this process is difficult to study in well-differentiated sex chromosomes. An earlier comparison between *A. franciscana* and *A. sinica* suggested that younger strata were acquired independently in the 2 lineages, which would make Artemia an ideal model for studying this stepwise process, but the fragmented assembly of *A. franciscana* limited this analysis (Elkrewi et al.2022). Furthermore, the Artemia genus includes multiple obligate parthenogenetic populations (Abatzopoulos 2018), and the Z chromosome has been implicated in their origin (Elkrewi et al. 2022). In this report, we present a chromosome-level genome assembly for *A. franciscana*, adding a valuable resource to the limited number of anostracan genomes, and use it to characterize the sex chromosome pair at the genomic and gene expression levels.

Results and Discussion

A Chromosome-Level Genome Assembly

We generated 5,006,105 PacBio circular consensus reads (CCS). Since an assembly of all the reads did not yield resolved Z and W haplotypes, we used female-specific kmers, generated using a k-mer subtraction approach with female and male short reads (Carvalho and Clark 2013; Elkrewi et al. 2021), to remove CCS reads originating from the W chromosome (25,784 reads, 0.52%; supplementary fig. S1). This was performed to avoid chimeric Z and W assemblies in the regions that retain some homology. The remaining 4,980,321 reads were assembled into 12,122 contigs using Hifiasm (Cheng et al. 2021). After removing 7,335 contigs representing alternative haplotypes, we scaffolded the assembly using evidence from CCS reads, RNA-seq reads, and previously published genomic mate pairs, resulting in 3,477 scaffolds, with an N50 of 590 KB (supplementary fig. S2).

We used a published high-density linkage map (Han et al. 2021) to anchor the scaffolds into 21 linkage groups. To improve the contiguity of the differentiated part of the Z chromosome (Huylmans et al. 2019), we performed coverage analysis to identify the scaffolds originating from the differentiated region of the Z chromosome (supplementary fig. S3) and anchored them using the linkage group 6 (LG6) (Z chromosome in the linkage map) markers. We then added the anchored differentiated region to the rest of the assembly and used the complete linkage map to assign scaffolds to their respective linkage groups (supplementary fig. S4). The resulting assembly was polished (gap filling and correction) using both the filtered CCS reads and male genomic short reads. The putative W reads (removed in the first step) were assembled separately (supplementary fig. S5), resulting in 506 putative W sequences, which were added to the assembly for the downstream analysis.

The final assembly has 2,118 scaffolds and an N50 of 43 MB, with 81% of the assembly in the 21 linkage groups (supplementary table S1). We ran BUSCO to assess the completeness of the genome, and 88.5% of BUSCOs were assembled completely, with <7% missing (Fig. 1a). We also checked for contamination using BlobTools, and the results show that there is <1% bacterial contamination and most sequences map to Arthropoda (Fig. 1b).

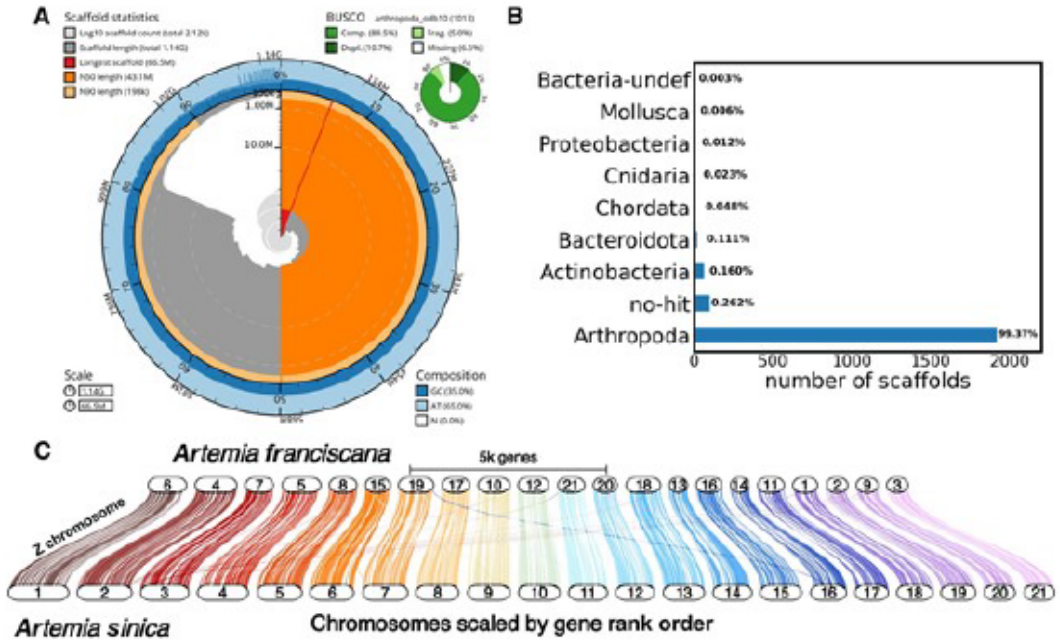


Fig. 1. Genome assembly and synteny. a) Genome assembly statistics, GC content, and BUSCO score using the arthropoda_odb10 data set. b) Barplot showing the number of sequences that map to the different phyla in the nt Blast database and the percentage of the total assembly length they represent. c) Synteny between the *A. franciscana* and *A. sinica* genomes.

As an additional quality check, since misassemblies cannot be ruled out given that anchoring of scaffolds was based on a linkage map rather than Hi-C data, we compared our assembly against the *A. sinica* chromosomes. Both the *A. franciscana* and *A. sinica* genomes were annotated using RNA and protein evidence (as described in the Repeat Content Characterization, Genome Annotation, and Synteny between the Artemia Genomes section). The annotations were used to examine and visualize the synteny between the 2 genomes using GENESPACE (Lovell et al. 2022). As Fig. 1c shows the genomes are highly syntenic, with no evidence for any large-scale rearrangements.

History and Extent of Differentiation of the Sex Chromosomes

Earlier work suggested the ZW pair of *A. franciscana* has a small but well-differentiated region, which no longer exists on the W, and a nonrecombining but undifferentiated region (Huylmans et al. 2019). We estimated the coverage across the genome using short-read male and female DNA in windows of 10,000 bp (supplementary fig. S6) and then used the ratio of female-to-male coverage to identify regions that have become well differentiated between the Z (LG6) and W chromosomes (Fig. 2a). We performed the analysis once with the W scaffolds included in the assembly and once without them. This makes it possible to identify regions that still share some sequence similarity between the Z and the W, as the W reads originating from those regions will map to the Z chromosome when the W is not included. On the other hand, regions that lost homology completely will have consistent low coverage regardless of whether the W scaffolds are included or not. A 13 MB region of the Z chromosome has coverage in females that is half the male coverage in both analyses. This is in agreement with the results obtained with the *A. franciscana* scaffold-level assembly anchored to the *A. sinica* genome (Elkrewi et al. 2022), but with much greater contiguity of the differentiated region. A smaller region adjacent to it shows a full reduction in female coverage when W scaffolds are included and

only partial reduction when they are not, a first line of evidence that it corresponds to a more recent and only moderately differentiated region.

ZW regions with limited differentiation, which show no or only partial coverage differences, can be detected by the presence of genetic variants specific to the W and therefore to females. We used pooled male and female RNA-seq libraries (Huylmans et al. 2019) to estimate the female:male F_{ST}, a measure of genetic differentiation, across the genome (supplementary fig. S7). We performed the analysis once with the W scaffolds included in the assembly and once without them. The analysis without the W scaffolds shows elevation in the male:female F_{ST} on the 2 sides of the differentiated region (Fig. 2b). The high F_{ST} is less pronounced on both sides when the W scaffolds are included, as the W-derived RNA reads preferentially map to the W. The decrease is more noticeable on the left side, suggesting a higher degree of differentiation. To further explore this, we estimated the median rate of synonymous substitutions (dS) between the transcripts on the putative W scaffolds which had female-specific expression patterns and their Z homologs for the different regions. We used dS, coverage, and F_{ST} patterns to define 3 strata: the ancestral S0, which shows high dS and consistent low female coverage estimates with and without the W scaffolds; S1, which shows intermediate coverage patterns and elevated F_{ST} when the W scaffolds are not included, suggesting the W reads carry many female-specific variants but still map to the region; and S2, which has elevated F_{ST} in the 2 cases and very low dS. Figure 2a shows the correspondence between those strata and the *A. sinica* strata and where the Z homologs of the identified W transcripts map on the 2 chromosomes. The ZW pairs in *A. franciscana* S0 map to the S1a stratum of *A. sinica*, suggesting that they might not be ancestral.

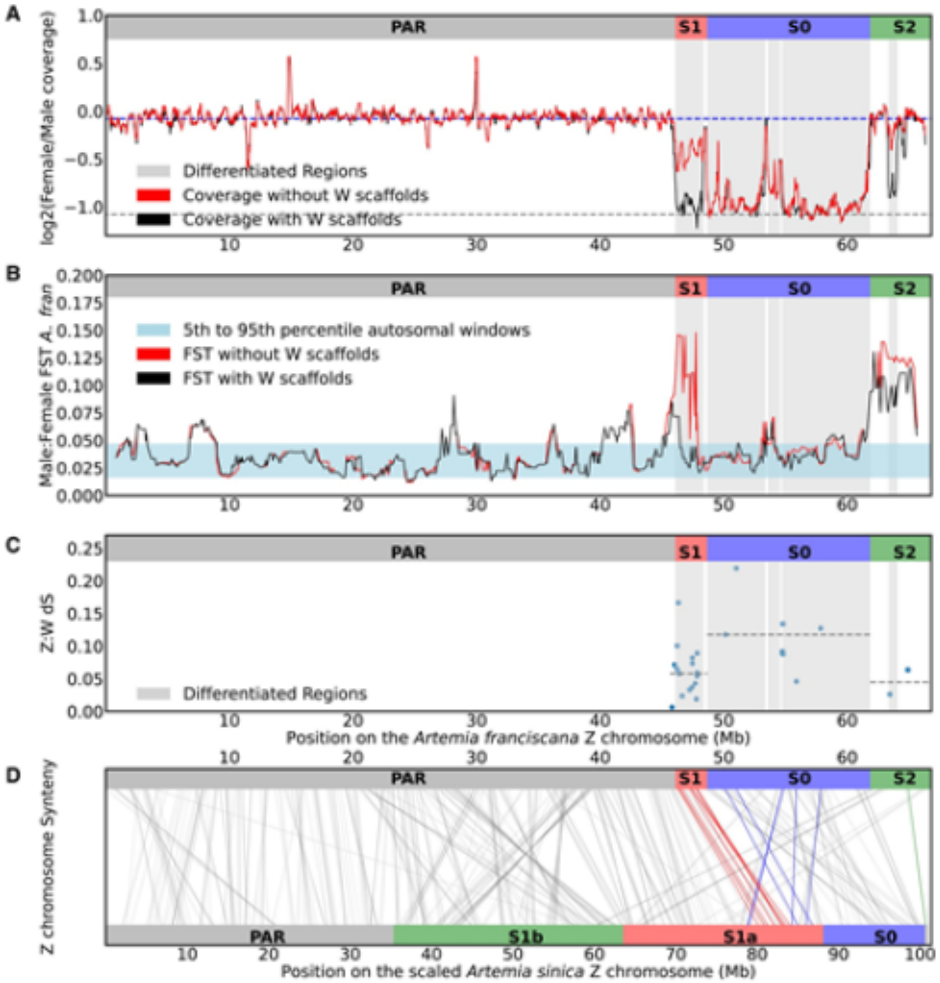


Fig. 2. Evolutionary strata of the ZW pair. a) Log2(female/male) coverage patterns of the Z chromosome. The rolling medians of the coverage for 30 consecutive (10,000 bp) windows, estimated with and without including the W scaffolds, are shown in the figure. The vertical shading highlights the differentiated region of the Z chromosomes. The top dashed line is the autosomal median for the analysis without W scaffolds, and the bottom dashed line is at median −1. b) The rolling medians of male:female FST per gene for 10 genes, estimated with and without including the W scaffolds, are shown. The horizontally shaded area covers the region between the 5th and 95th percentiles for autosomal windows with the W included. c) dS values between the W transcripts and their Z homologs. The dashed lines correspond to the median of the dS values in the different strata. d) Synteny between the *A. sinica* and *A. franciscana* Z chromosomes highlighting the different strata (inferred here for *A. franciscana* and found in Elkrewi et al 2022 for *A. sinica*): S0, S1, and S2. The lines connect the locations of the reciprocal best hits on the chromosomes, with the Z homologs of the identified W transcripts colored according to their strata in *A. franciscana*.

Full Dosage Compensation and Repeat Composition on the ZW Pair

We estimated levels of expression for all annotated genes from published male and female head and gonad RNA-seq data (Huylmans et al. 2019). Gene expression does not differ between the differentiated region of the Z and the autosomes in either male and female heads or gonads (Fig. 3, $P > 0.05$ with Wilcoxon rank sum tests). This is consistent with dosage compensation, i.e. a mechanism to balance the expression of the sex chromosomes and autosomes in both sexes in species with differentiated sex chromosomes, as reported in earlier work (Huylmans et al. 2019). Most ZW systems that have been studied so far, such as snakes and birds, seem to lack a chromosome-wide mechanism of dosage compensation (Gu and Walters 2017). So far, Lepidoptera (moths and butterflies) have been the only clear exception to this rule (Gu and Walters 2017). Our confirmation that chromosome-wide compensation also occurs in *Artemia* makes it a promising model for understanding why and how such compensatory mechanisms arise in female heterogametic species.

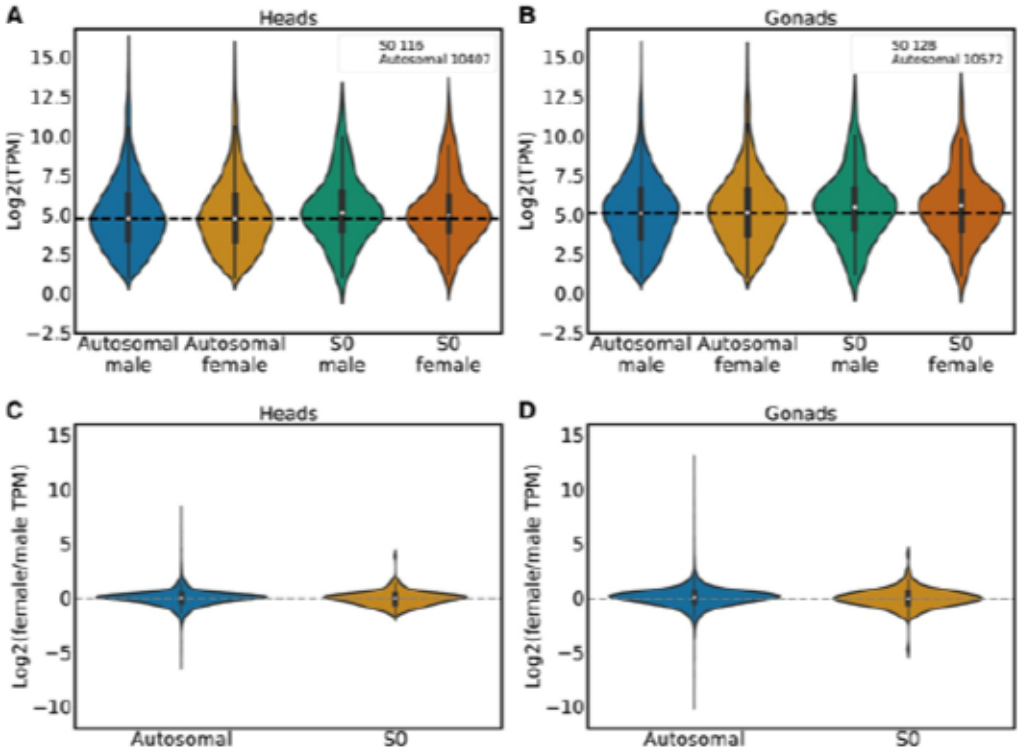


Fig. 3. Dosage compensation. a) The Log2 of the expression of autosomal and the Z differentiated region genes in male and female heads. The legend shows the number of genes used in the analysis. b) The Log2 expression of autosomal and the Z differentiated region (S0) genes in gonads. The legend shows the number of genes used in the analysis. c) The Log2 of female/male expression for the autosomal and Z differentiated genes in heads. d) The Log2 of female/male expression for the autosomal and Z differentiated genes in gonads. Only genes with TPM $\geq$ 0.5 in their respective male and female tissues were used in these analyses.

Finally, the absence of recombination between sex chromosomes often leads to the accumulation of transposable elements (TEs) and other repetitive sequences on the Y or W chromosome (Dechaud et al. 2019). Approximately 2/3 of the genome in *A. franciscana* are covered by repetitive elements (supplementary table S2). Retroelements account for 43% of the repeat content in this species. We did not observe striking differences in the overall repeat content between W scaffolds and either S0 region or autosomes. However, W scaffolds that still have homologs to the S0 region, and which should correspond to the most differentiated part of this chromosome, have more retroelements (48%) in relation to those in S0 region, autosomes, and W scaffolds (supplementary table S2). Overall, even this region of the W only shows a modest enrichment in repeats compared with the autosomes (67% vs. 66%) and none relative to the S0 region of the Z (68%). While it is possible that this pattern reflects a slow accumulation of repeats in the nonrecombining part of the W in this clade (or limited power to resolve repeats in the assembly), another possibility is that the ancestral nonrecombining region of the W has been lost entirely and that all regions studied here are relatively young. Complete loss of the original W-specific region would also account for the lack of ZW homologs that map to the S0 of both *A. franciscana* and *A. sinica* and potentially explain why a global mechanism of dosage compensation was selected for in this lineage.

Materials and Methods

DNA Extraction and Sequencing

High molecular weight DNA was extracted from 2 unmated females from the great salt lake, purchased from Sanders (Utah, United States), with the Qiagen Genomic-Tip 20/G Kit, and sequenced on a PacBio Sequel II SMRT cell at the Vienna Biocenter sequencing facility. The number of females was chosen to ensure enough genetic material and also limit the amount of genetic variability, which would complicate the assembly.

Genome Assembly

The consensus sequences of the subreads in the raw bam file were generated using the PacBio ccs tool (option -all) (v6.4.0, https://github.com/PacificBiosciences/ccs). The bbduk.sh script (from BBmap) was used to identify female-specific 21-mers from *A. franciscana* male (SRR19741748) and female short reads (SRR19741747), and the resulting kmers were used to remove putative W-specific CCS reads (reads with 20% or more female-specific kmers, 0.2 mkf) (Bushnell 2014). The filtered CCS reads were then assembled using Hifiasm (--hg-size 1g --n-hap 4 -r 5 -s 0.7 -N 150) (v0.19.4, Cheng et al. 2021). The primary assembly was then purged of duplicates with female short reads using purged_dups (v1.2.5, Guan et al. 2020). We scaffolded the purged assembly with the filtered CCS reads using LongStitch (ntlink + arcs, v1.0.4, Coombe et al. 2021). We independently scaffolded the assembly with male RNA-seq reads (from Huylmans et al. 2019) using Rascaf (Song et al. 2016). We then mapped the 2 resulting assemblies to the input assembly with minimap2 (Li 2018) and used a python script to identify the merges that are supported by both approaches (supplementary fig. S3). RagTag (v2.1.0, Alonge et al. 2022) was used to implement those merges. Redundans (v0.14a, Pryszcz and Gabaldón 2016) was then used with long insert mate pairs (SRR6980924, 8 MB) to further scaffold the assembly.

Scaffolding Using the Published Linkage Map

The published *A. franciscana* linkage map was used to scaffold the draft assembly into linkage groups (Han et al. 2021). The SLAF markers were mapped to the assembly using BWA-MEM (v0.7.17-r1198-dirty, Li 2013), and the alignments, along with the linkage map, were used by Chromonomer (v1.14, Catchen et al. 2020) to anchor the scaffolds. The scaffolding was done in 2 steps, with the first to anchor the differentiated region of the Z chromosome. To do that, we estimated the coverage for the assembled scaffolds as described in the coverage section and selected putative Z-specific scaffolds (114 scaffolds). We used the perl script BreakScaffolds (https://github.com/aubombarely/GenoToolBox) to split the scaffolds at stretches > 100 Ns, which resulted in 126 scaffolds. The LG6 markers were mapped to those scaffolds, and the LG6 linkage map was used for anchoring (54 scaffolds were anchored into a 13.5 MB region). The output was added to the remaining scaffolds, and the complete linkage map was used to anchor everything into the 21 linkage groups. In order to avoid splitting the differentiated region, we ran 2 iterations of Chromonomer. The first iteration was run with the option "--disable_splitting" to identify the best location for the scaffolded differentiated region. We then modified the original linkage map to retain only the differentiated region markers and location appearing in the output file (CHRR_linkage_map.tsv) from the first iteration. The second iteration was run with the modified linkage map with the option "—rescaffold." This ensured that Chromonomer was able to break and rescaffold regions with inconsistent markers without splitting the differentiated region.

Polishing and Quality Assessment

TGS-GapCloser (v1.1.1, Xu et al. 2020) was used to fill the gaps in the assembly using the filtered CCS reads. The first round of polishing with the filtered CCS reads and male short reads used Racon (v1.6.0, Vaser et al. 2017) and Merfin (v1.1, Formenti et al. 2022) through the automated-polishing.sh script (from https://github.com/arangrhie/T2T-Polish/tree/master/automated_polishing, modified from Mc Cartney et al. 2022), and the second round used nextpolish2 (v0.1.0-758ef0a, Hu et al. 2024). The assessment of the assembly completeness was done using BUSCO (v5.2.2), with the arthropoda_odb10 data set (Manni et al. 2021), and BlobTools was used to assess and visualize the level of contamination and genome statistics (Laetsch and Blaxter 2017).

Genomic Coverage, F_{ST} Analysis, and Identification of the Differentiated Region

For the coverage analysis, male and female Illumina genomic short reads (from Elkrewi et al. 2022) were mapped to the genome using Bowtie2 (--end-to-end --sensitive) (v2.4.4, Langmead and Salzberg 2012). The resulting SAM files were filtered for unique alignments using (grep -vw "XS:i"), and the coverage was estimated for windows of 10,000 bp using soap.coverage (version 2.7.7, https://github.com/gigascience/bgi-soap2/tree/master/tools/soap.coverage/2.7.7).

In the Fst analysis, head and gonad RNA-seq samples from 10 males and 10 females of *A. franciscana* (from Huylmans et al. 2019) were pooled by sex and then mapped to the genome using STAR (v2.7.9a, Dobin et al. 2013). The SAMtools mpileup command was used to generate a text pileup output from the 2 sorted alignment bam files (v1.18, Li et al. 2009). Grenedalf (v0.2.0, Czech et al. 2023) sync was used to get a sync file, which was used as input to the PoPoolation2 perl scripts (create-genewise-sync.pl and fst-sliding.pl) along with the GTF file (Repeat Content Characterization, Genome Annotation, and Synteny between the Artemia Genomes section) to estimate the Fst per gene (Kofler et al. 2011).

The coordinates for the differentiated region (gray-shaded areas; Fig. 2a) were defined using the approach described in Elkrewi et al. (2022), as regions where the Log2(female/male coverage) drops below the median of the Log2(female/male coverage) of autosomal windows (−0.5), and as long as the coverage does not go above the defined threshold, the extension of the region continues. The resulting coordinates were (46,085,001 to 48,385,001), (48,665,001 to 53,365,001), (53,585,001 to 54,575,001), (54,715,001 to 61,855,001), and (63,345,001 to 64,075,001), shaded in gray in Fig. 2a–c.

The annotated W transcripts were mapped to the rest of the transcriptome using BlastN (Altschul et al. 1990), and the reciprocal best hits were identified as the homologs using a customized perl script. W transcripts with homologs on the Z chromosome and the sum of head and gonad female/(male + female) expression ≥ 0.9 were used for estimating the rate of synonymous substitution. The ZW homologs were aligned with the TranslatorX package (Abascal et al. 2010) with the "gblocks" option to filter out unreliable sections of the alignment. The dN and dS values were obtained with KaKs_calculator2.0 (Wang et al. 2010) using the Nei and Gojobori algorithm (NG). Alignments shorter than 300 bp were excluded from the analysis.

Repeat Content Characterization, Genome Annotation, and Synteny between the *Artemia* Genomes

RepeatModeler (v2.0.4, Flynn et al. 2020) was applied on the *A. franciscana* genome assembly to generate a de novo library of repeat families. The sequences of unknown TEs were classified using DeepTE (Yan et al. 2020) with the options "-sp M -m M." Subsequently, Class II TEs (DNA transposons) were categorized further into superfamilies and this was done with the following parameters "sp M -m M -fam ClassII." These were then combined with both Class I known repeat libraries RepeatModeler and DeepTE and used for the characterization of repeat content and masking of the genome by RepeatMasker (v4.1.5, Tarailo-Graovac and Chen 2009). Evidence from both RNA and protein was used for the annotation of predicted genes on the soft-masked genome using BRAKERv2.1.6 (Hoff et al. 2019). For RNA, reads were aligned to the genome using STAR (v2.7.9a, Dobin et al. 2013) in "--twopassMode" and sorted with SAMtools v1.18 (Li et al. 2009). To generate protein hints, all arthropoda protein sequences were downloaded from OrthoDBv11 (https://www.orthodb.org) and then concatenated into a single fasta that was then aligned to the soft-masked genome using ProHint (Brůna et al. 2020) to give predicted protein location in the genome in the form of gff3. BRAKER2, automated gene prediction based on successive runs of GenemarkEP+ and Augustus, was applied on protein hints and sorted RNA alignments with options "–etpmode; –softmasking;." BRAKER2 was run twice with the second round being performed as before but with an additional hint file generated from the first gene prediction run. AGAT (Dainat et al. 2023) was then used to remove isoforms and incomplete genes without start and/or stop codons and filtered out those with <100 bp length of open reading frames. Protein sequences and coding sequences were extracted using getAnnoFastaFromJoingenes.py, and their completeness was assessed with BUSCO (v5.2.2).

An annotation was produced for the *A. sinica* genome using the same approach described above, and the overall synteny was examined and visualized using GENESPACE v. 0.94 (Lovell et al. 2022). Additionally, the annotated protein sequences for *A. sinica* and *A. franciscana* were mapped to each other using pblat (Wang and Kong 2019) (protein target and query) and reciprocal best hits were found using a customized python script and used for producing Fig. 2d.

Expression Analysis

Illumina RNA reads of 2 biological replicates of heads and gonads of each sex were mapped to the genome using STAR (v2.7.9a, Dobin et al. 2013) with the following parameters "--twopassMode basic --quantMode TranscriptomeSAM GeneCounts" and additional option --quantTranscriptomeBan IndelSoftclipSingleend to generate bam alignments that are acceptable as inputs to RSEM (Li and Dewey 2011). Transcript abundances of genes (in TPM) were estimated using rsem-calculate-expression in RSEM with options "--paired-end --alignments --estimate-rspd --strandedness reverse". We then used NormalyzerDE (Willforss et al. 2019) to perform quantile normalization across samples for each tissue separately. We applied different cutoffs of TPM $\geq$ 0.0, TPM $\geq$ 0.5, and TPM $\geq$ 1, in heads and gonads separately, to assess dosage compensation, whereby the genes with average replicate expression above the thresholds were retained in both sexes. The average TPM values for each tissue in each sex were then normalized with Log2 (Fig. 2a–d, supplementary figs. S8a and b and S9a and b).
The assembly pipeline and the scripts used in the analysis can be accessed on the gitpage (https://github.com/Melkrewi/Artemia_franciscana_genome/). The raw data are available on the National Center for Biotechnology Information (NCBI) short-read archive (BioProject

number PRJNA1017357). The final assembly, the annotation, and the supplementary data sets are available on (https://doi.org/10.15479/AT:ISTA:14705). All genomic and transcriptomic samples used for this study and the steps of the analysis they were used in are described in supplementary table S3.

References

Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010:38(suppl_2):W7–W13. https://doi.org/10.1093/nar/gkq291.

Abatzopoulos TJ. The repeated emergence of asexuality, the hidden genomes and the role of parthenogenetic rare males in the brine shrimp Artemia. J Biol Res-Thessalon. 2018:25(1):7. https://doi.org/10.1186/s40709-018-0078-2.

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022:23(1):258. https://doi.org/10.1186/s13059-022-02823-7.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Baxevanis AD, Kappas I, Abatzopoulos TJ. Molecular phylogenetics and asexuality in the brine shrimp Artemia. Mol Phylogenet Evol. 2006:40(3):724–738. https://doi.org/10.1016/j.ympev.2006.04.010.

Bowen ST. The genetics of Artemia salina. III. Effects of X-irradiation and of freezing upon cysts. Biol Bull. 1963:125(3):431–440. https://doi.org/10.2307/1539357.

Boyer L, Jabbour-Zahab R, Joncour P, Glémin S, Haag CR, Lenormand T. Asexual male production by ZW recombination in Artemia parthenogenetica. Evolution 2023:77(1):1–12. https://doi.org/10.1093/evolut/qpac008.

Brů na T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic geneprediction with self-training in the space of genes and proteins. NAR Genomics Bioinforma. 2020:2(2):lqaa026. https://doi.org/10.1093/nargab/lqaa026.

Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Berkeley, CA (United States): Lawrence Berkeley National Lab. (LBNL); 2014.https://www.osti.gov/biblio/1241166 (Accessed 2023 September 4).

Carvalho AB, Clark AG. Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Res. 2013:23(11):1894–1907. https://doi.org/10.1101/gr.156034.113.

Castellucci F, Luchetti A, Mantovani B. Exploring mitogenome evolution in Branchiopoda (Crustacea) lineages reveals gene order rearrangements in Cladocera. Sci Rep. 2022:12(1):4931. https://doi.org/10.1038/s41598-022-08873-y.

Catchen J, Amores A, Bassham S. Chromonomer: a tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. G3 GenesGenomesGenetics. 2020:10(11):4115–4128. https://doi.org/10.1534/g3.120.401485.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18(2):170–175. https://doi.org/10.1038/s41592-020-01056-5.

Coombe L, Li JX, Lo T, Wong J, Nikolic V, Warren RL, Birol I. LongStitch: high-quality genome assembly correction and scaffolding using long reads. BMC Bioinformatics. 2021:22(1):534. https://doi.org/10.1186/s12859-021-04451-7.

Criel G, Macrae T. Reproductive biology of Artemia. In: Artemia: basic and applied biology. Kluwer Academic; 2002. p. 39–128. http://hdl.handle.net/1854/LU-150279 (Accessed 2023 September 4).
Czech L, Spence JP, Expósito-Alonso M. grenedalf: population genetic statistics for the next generation of pool sequencing. arXiv. 2023. https://doi.org/10.48550/ARXIV.2306.11622.
Dainat J, Hereñú D, Murray DKD, Davis E, Crouch K, LucileSol, Agostinho N, pascal-git, Zollman Z, tayyrov. NBISweden/AGAT: AGAT-v1.2.0 2023. https://doi.org/10.5281/ZENODO.8178877.
Dechaud C, Volff J-N, Schartl M, Naville M. Sex and the TEs: transposable elements in sexual development and function in animals. Mob DNA. 2019:10(1):1–15. https://doi.org/10.1186/s13100-019-0185-0.
De Vos S, Bossier P, Van Stappen G, Vercauteren I, Sorgeloos P, Vuylsteke M. A first AFLP-based genetic linkage map for brine shrimp Artemia franciscana and its application in mapping the sex locus. PLoS One 2013:8(3):e57585. https://doi.org/10.1371/journal.pone.0057585.
De Vos S, Rombauts S, Coussement L, Dermauw W, Vuylsteke M,Sorgeloos P, Clegg JS, Nambu Z, Van Nieuwerburgh F, Norouzitallab P, et al. The genome of the extremophile Artemia provides insight into strategies to cope with extreme environments. BMC Genomics. 2021:22(1):635. https://doi.org/10.1186/s12864-021-07937-z.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.
Eimanifar A, Van Stappen G, Wink M. Geographical distribution and evolutionary divergence times of Asian populations of the brine shrimp Artemia (Crustacea, Anostraca): divergence times of Asian Artemia. Zool J Linn Soc. 2015:174(3):447–458. https://doi.org/10.1111/zoj.12242.
Elkrewi M, Khauratovich U, Toups MA, Bett VK, Mrnjavac A, Macon A, Fraisse C, Sax L, Huylmans AK, Hontoria F, et al. ZW sex chromosome evolution and contagious parthenogenesis in Artemia brine shrimp. Genetics 2022:222(2):iyac123. https://doi.org/10.1093/genetics/iyac123.
Elkrewi M, Moldovan MA, Picard MAL, Vicoso B. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. Mol Biol Evol. 2021:38(12):5345–5358. https://doi.org/10.1093/molbev/msab178.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117(17):9451–9457. https://doi.org/10.1073/pnas.1921046117.
Formenti G, Rhie A, Walenz BP, Thibaud-Nissen F, Shafin K, Koren S, Myers EW, Jarvis ED, Phillippy AM. Merfin: improved variant filtering, assembly evaluation and polishing via k-mer validation. Nat Methods. 2022:19(6):696–704. https://doi.org/10.1038/s41592-022-01445-y.
Gu L, Walters JR. Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details. Genome Biol Evol. 2017:9(9):2461–2476. https://doi.org/10.1093/gbe/evx154.
Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020:36(9):2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.
Han X, Ren Y, Ouyang X, Zhang B, Sui L. Construction of a high-density genetic linkage map and QTL mapping for sex and growth traits in Artemia franciscana. Aquaculture 2021:540:736692. https://doi.org/10.1016/j.aquaculture.2021.736692.

Handley L-JL, Ceplitis H, Ellegren H. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics 2004:167(1):367–376. https://doi.org/10.1534/genetics.167.1.367.
Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Methods Mol Biol. 2019:1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.
Hu J, Wang Z, Liang F, Liu S, Ye K, Wang D-P. NextPolish2: a repeat aware polishing tool for genomes assembled using HiFi long reads. Genomics, Proteomics & Bioinformatics. 2024:qzad009. https://doi.org/10.1093/gpbjnl/qzad009
Huylmans AK, Toups MA, Macon A, Gammerdinger WJ, Vicoso B. Sex-biased gene expression and dosage compensation on the Artemia franciscana Z-chromosome. Genome Biol Evol. 2019:11(4):1033–1044. https://doi.org/10.1093/gbe/evz053.
Jo E, Lee SJ, Choi E, Kim J, Lee SG, Lee JH, Kim J-H, Park H. Whole genome survey and microsatellite motif identification of Artemia franciscana. Biosci Rep. 2021:41(3):BSR20203868. https://doi.org/10.1042/BSR20203868.
Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 2011:6(1):e15925. https://doi.org/10.1371/journal.pone.0015925.
Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Res. 2017:6:1287. https://doi.org/10.12688/f1000research.12232.1.
Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. Science 1999:286(5441):964–967. https://doi.org/10.1126/science.286.5441.964.
Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012:9(4):357–359. https://doi.org/10.1038/nmeth.1923.
Lavens P, Sorgeloos P. The history, present status and prospects of the availability of Artemia cysts for aquaculture. Aquaculture 2000: 181(3-4):397–403. https://doi.org/10.1016/S0044-8486(99)00233-1.
Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. https://doi.org/10.48550/ARXIV.1303.3997.
Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics. 2011:12(1):323. https://doi.org/10.1186/1471-2105-12-323.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife 2022:11:e78526. https://doi.org/10.7554/eLife.78526.
Maniatsi S, Kappas I, Baxevanis A, Farmaki T, Abatzopoulos T. Sharp phylogeographic breaks and patterns of genealogical concordance in the brine shrimp Artemia franciscana. Int J Mol Sci. 2009:10(12): 5455–5470. https://doi.org/10.3390/ijms10125455.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO up date: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021:38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.

Mc Cartney AM, Shafin K, Alonge M, Bzikadze AV, Formenti G, Fungtammasan A, Howe K, Jain C, Koren S, Logsdon GA, et al. Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies. Nat Methods. 2022:19(6): 687–695. https://doi.org/10.1038/s41592-022-01440-3.
Nunes BS, Carvalho FD, Guilhermino LM, Van Stappen G. Use of the genus Artemia in ecotoxicity testing. Environ Pollut. 2006:144(2): 453–462. https://doi.org/10.1016/j.envpol.2005.12.037.
Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 2016:44(12): e113–e113. https://doi.org/10.1093/nar/gkw294.
Rajabi S, Ramazani A, Hamidi M, Naji T. Artemia salina as a model organism in toxicity assessment of nanoparticles. DARU J Pharm Sci. 2015:23(1):20. https://doi.org/10.1186/s40199-015-0105-x.
Song L, Shankar DS, Florea L. Rascaf: improving genome assembly with RNA sequencing data. Plant Genome. 2016:9(3):plantgenome2016.03.0027. https://doi.org/10.3835/plantgenome2016.03.0027.
Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinforma. 2009:25:4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.
Van Stappen G, Sui L, Hoa VN, Tamtin M, Nyonje B, de Medeiros Rocha R, Sorgeloos P, Gajardo G. Review on integrated production of the brine shrimp Artemia in solar salt ponds. Rev Aquac. 2020:12(2):1054–1071. https://doi.org/10.1111/raq.12371.
Vaser R, Sovic´ I, Nagarajan N, Šikic´ M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017:27(5):737–746. https://doi.org/10.1101/gr.214270.116.
Wang M, Kong L. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinformatics. 2019: 20(1):1–4. https://doi.org/10.1186/s12859-019-2597-8.
Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. Kaks_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010:8(1):77–80. https://doi.org/10.1016/S1672-0229(10)60008-3.
Willforss J, Chawade A, Levander F. NormalyzerDE: online tool for improved normalization of omics expression data and high-sensitivity differential expression analysis. J Proteome Res. 2019:18(2): 732–740. https://doi.org/10.1021/acs.jproteome.8b00523.
Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, Fan G, Liu X, Xu X, Deng L, et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience 2020:9(9):giaa094.
https://doi.org/10.1093/gigascience/giaa094.
Yan H, Bombarely A, Li S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. Bioinformatics. 2020:36(15):4269–4275. https://doi.org/10.1093/bioinformatics/btaa519.

Supplementary Materials

Supplementary Datasets

Supplementary Dataset 1: *Artemia franciscana* genome assembly and annotation

Supplementary Dataset 2: *Artemia sinica* annotation

Supplementary Dataset 3: Male and Female genomic coverage

Supplementary Dataset 4: Male:female F_{ST} per gene

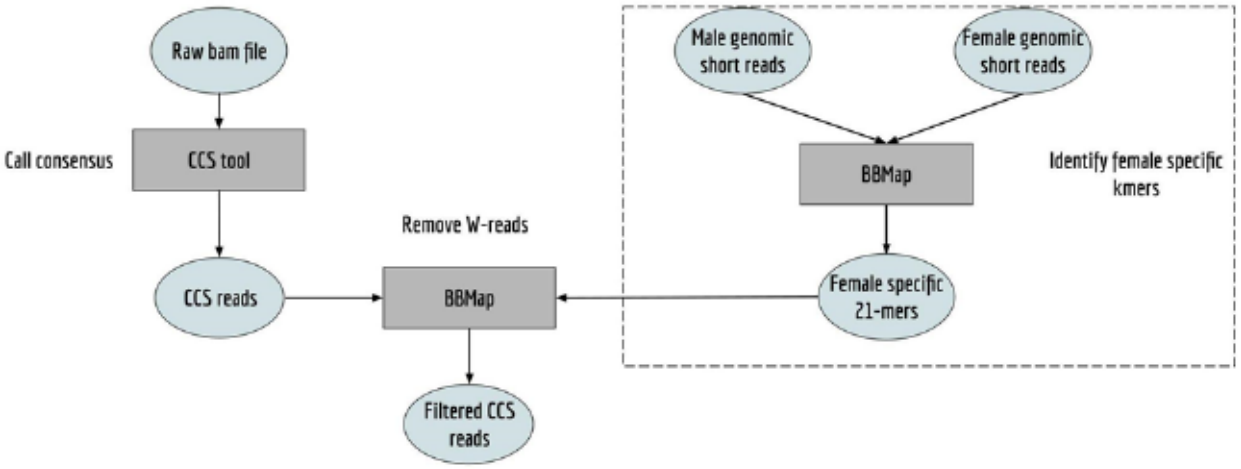
Supplementary Dataset 5: Normalized gene expression

Supplementary Dataset 6: W-specific sequences

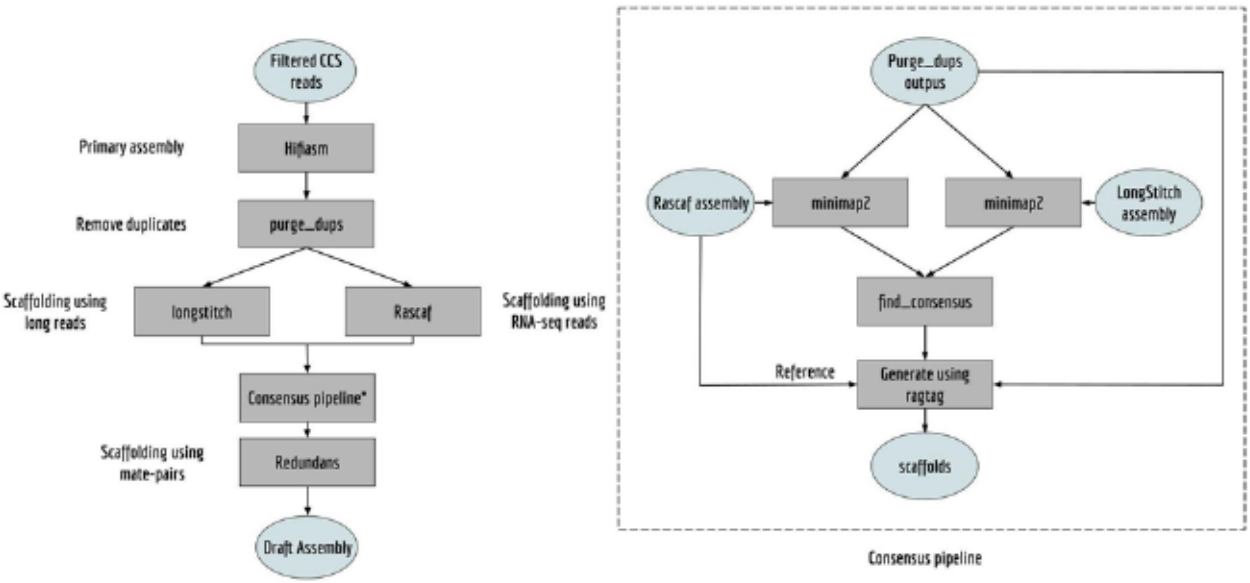
Supplementary Dataset 7: dN/dS results Z-W pairs in *A. franciscana*

All supplementary datasets can be downloaded using the following link: https://doi.org/10.15479/AT:ISTA:14705

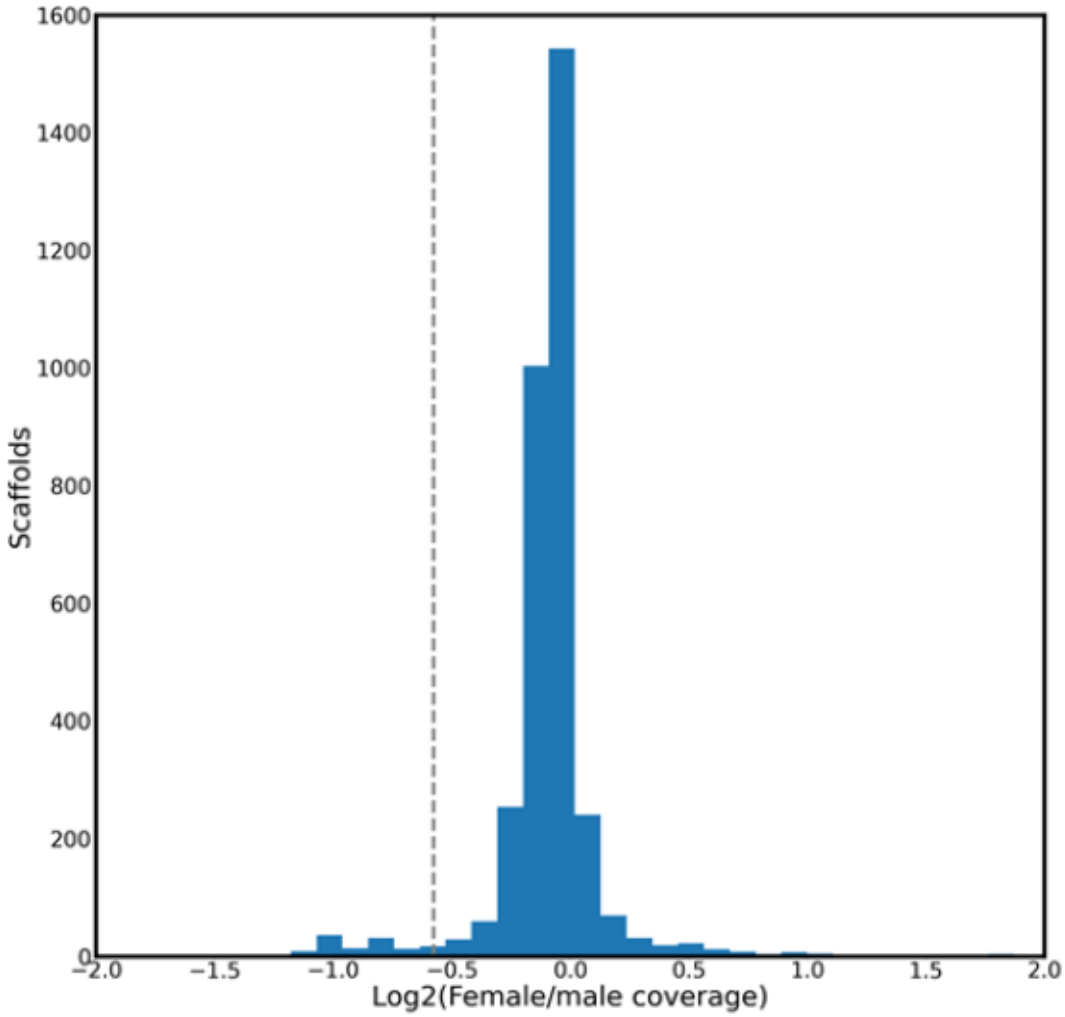
Supplementary Figures



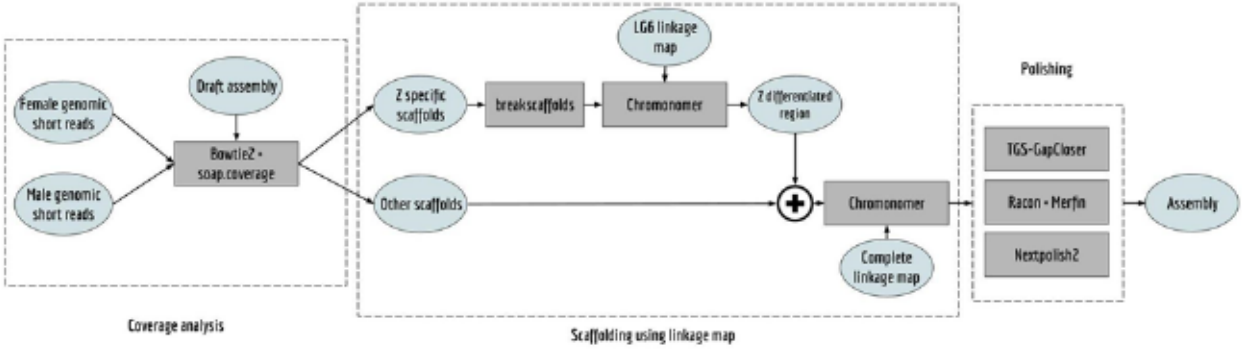
Supplementary Figure 1: the pipeline from generating the consensus reads, and Identifying and removing putative W-reads using female specific k-mers identified using a subtraction approach with male and female genomic short reads.



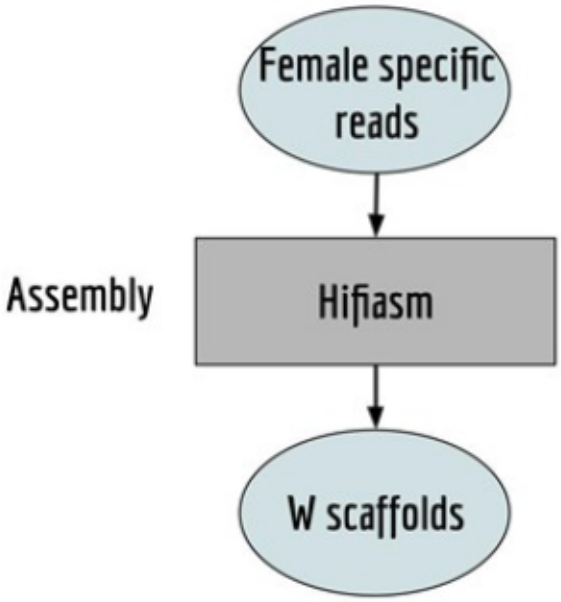
Supplementary Figure 2: Assembling the genome and scaffolding with genomic long reads and RNA-seq reads



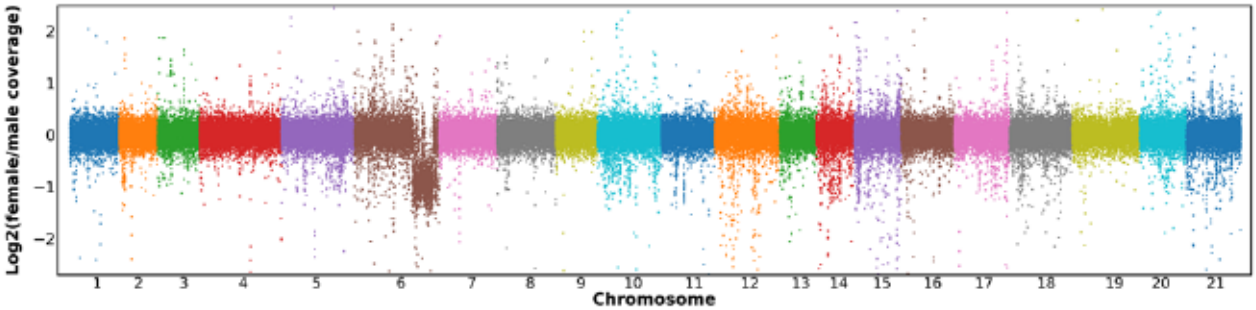
Supplementary Figure 3: Coverage analysis used to identify the scaffolds belonging to the differentiated region of the Z chromosome. The gray-dashed line represents the threshold used: (Log2(female/male) < median(Log2(female/male)-0.5).



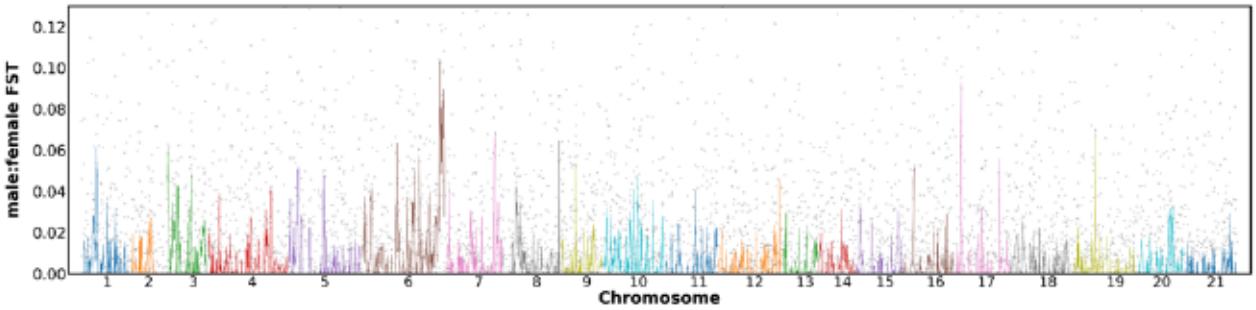
Supplementary Figure 4: The pipeline for scaffolding the genome using the published linkage map



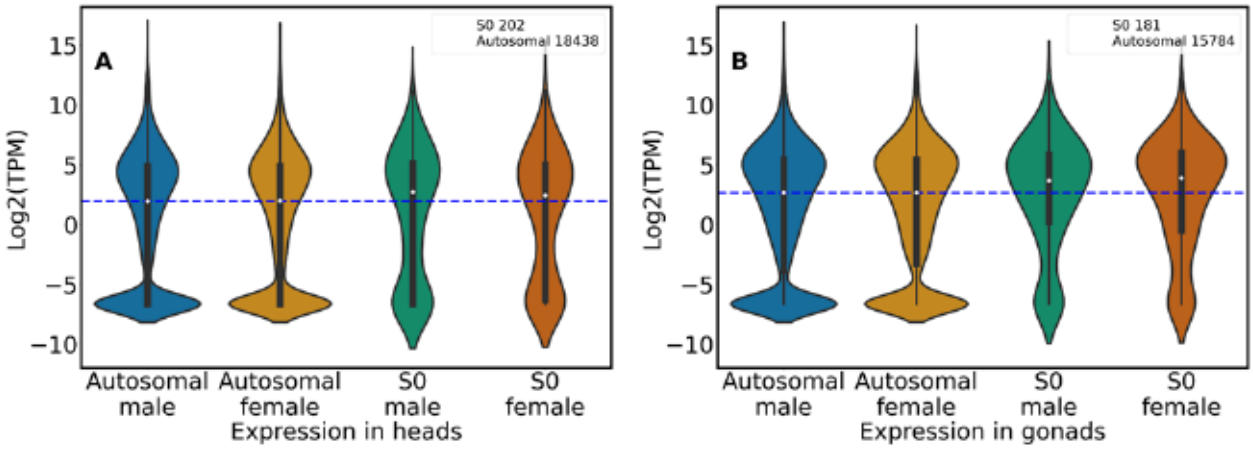
Supplementary Figure 5: Assembly of W scaffolds



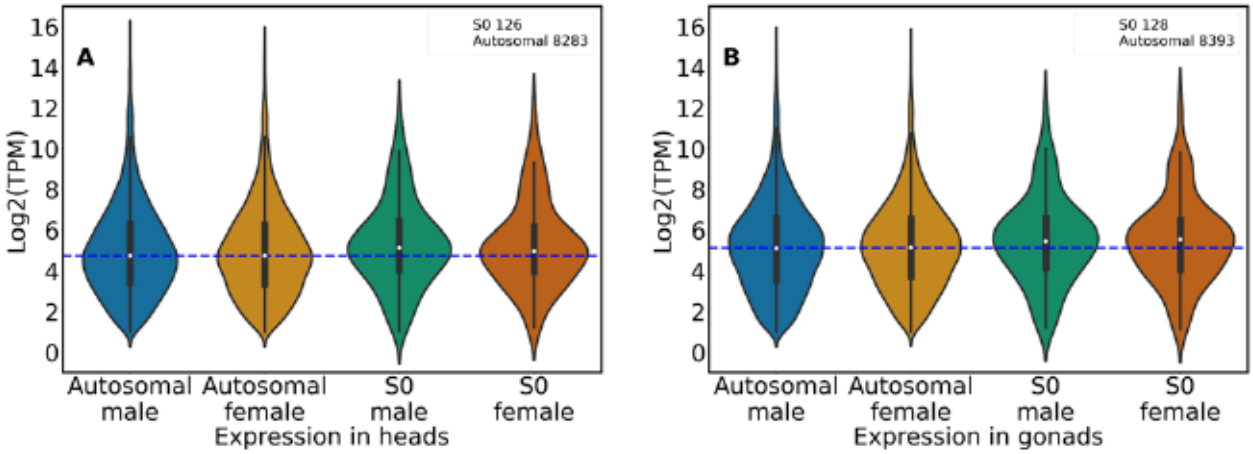
Supplementary Figure 6: Log2(Female/Male) coverage patterns across the whole genome (W-scaffolds included).



Supplementary Figure 7: male:female F_{ST} patterns across the whole genome (W-scaffolds included). The gray dots are the per gene F_{ST}, and the colored tracing is the rolling median of 10 consecutive genes.



Supplementary Figure 8: A) The Log2 of the expression of Autosomal and the Z differentiated region genes in male and female heads. The legend shows the number of genes used in the analysis (TPM >= 0 in males and females). B) The Log2 expression of Autosomal and the Z differentiated region genes in gonads. The legend shows the number of genes used in the analysis (TPM >= 0 in males and females).



Supplementary Figure 9: A) The Log2 of the expression of Autosomal and the Z differentiated region genes in male and female heads. The legend shows the number of genes used in the analysis (TPM >= 1 in males and females). B) The Log2 expression of Autosomal and the Z differentiated region genes in gonads. The legend shows the number of genes used in the analysis (TPM >= 1 in males and females).

Supplementary Tables

Supplementary Table 1: The statistics of the genome and transcriptome assemblies

Assembly	Genome Size (MB)	1135.764093
	Number of chromosomes	21
	Linkage map anchoring ratio	81%

	Number of scaffolds/contigs	2118
	Longest scaffold/contig (Mb)	66
	N50 (MB)	43
	Average length (MB)	0.54
	Gaps	2470
	N_count	520572
	GC content (%)	34.95%
Protein-coding genes	Number	19421
	Mean gene length (bp)	30503
	Exons/introns per gene	~5.2 exons per gene ~4.2 introns per gene
	Mean exon/intron length (bp)	~252 bp exons per gene ~6964 bp introns per gene
Busco of protein sequences of annotated genes	complete	90.2%
	single-copy	76.9%
	duplicated	13.3%
Repeats	Size (Mb)	749 (65.94%)
	DNA transposons (MB)	45 (3.98%)
	SINEs (MB)	9.2 (0.81%)

	LINEs (MB)	134 (11.78%)
	LTRs (MB)	351 (30.91 %)
	Unclassified transposons (Mb)	125 (11.08%)

Supplementary Table 2: Proportion of Transposable Elements in W scaffolds, SO region, Autosomes, W-scaffolds that have homologs to S0 region and across whole genome of *A. franciscana*

Type	**Whole genome (bp)**	**Autosomes (bp)**	**ChrZ S0 region (bp)**	**W scaffolds (bp)**	**W scaffolds with S0 homolog (bp)**
total length	1135764093	1101127100	13110000	21526993	819737
bases masked	748899995 (65.94 %)	726202062 (65.95%)	8967129 (68.40%)	13731055 (63.79 %)	553123 (67.48 %)
Retroelements	494046611 (43.50 %)	479334586 (43.53 %)	5500604 (41.96%)	9211377 (42.79 %)	396462 (48.36 %)
DNA transposons	5228968 (3.98 %)	43620029 (3.96 %)	615807 (4.70 %)	993357 (4.61 %)	39439 (4.81 %)
Rolling-circles	59988424 (5.28 %)	58336423 (5.30 %)	1013560 (7.73 %)	638620 (2.97 %)	15105 (1.84 %)
Unclassified	125827194 (11.08 %)	121790622 (11.06 %)	1560033 (11.90 %)	2475977 (11.50 %)	89778 (10.95 %)

Supplementary Table 3: Genomic and transcriptomic samples used for this study and what steps of the analysis they were used in. This table is available here: https://github.com/Melkrewi/Artemia_franciscana_genome/blob/048557e7ed5fd57b069516f5c3da652ad9d4c1cf/Artemia_francisca_genome_list_of_samples.xlsx

Chapter 3: Chromatin Landscape Is Associated With Sex-Biased Expression and Drosophila-Like Dosage Compensation of the Z Chromosome in *Artemia franciscana*

Vincent Kiplangat Bett [1*], Minerva Susana Trejo-Arellano [1], Beatriz Vicoso [1*]

Abstract

The males and females of the brine shrimp *Artemia franciscana* are highly dimorphic, and this dimorphism is associated with substantial sex-biased gene expression in heads and gonads. How these sex-specific patterns of expression are regulated at the molecular level is unknown. *A. franciscana* also has differentiated ZW sex chromosomes, with complete dosage compensation, but the molecular mechanism through which compensation is achieved is unknown. Here, we conducted CUT&TAG assays targeting 7 post-translational histone modifications (H3K27me3, H3K9me2, H3K9me3, H3K36me3, H3K27ac, H3K4me3, and H4K16ac) in heads and gonads of *A. franciscana*, allowing us to divide the genome into 12 chromatin states. We further defined functional chromatin signatures for all genes, which were correlated with transcript level abundances. Differences in the occupancy of the profiled epigenetic marks between sexes were associated with differential gene expression between males and females. Finally, we found a significant enrichment of the permissive H4K16ac histone mark in the Z-specific region in both tissues of females but not males, supporting the role of this histone mark in mediating dosage compensation of the Z chromosome.

Introduction

The males and females of many species differ in their morphology, physiology, and/or behavior. While the primary signal for sex determination is encoded on the sex chromosomes in species with genetic sex determination, sexual differentiation is thought to rely on differential gene expression (sex-biased genes) between males and females throughout the genome (Parsch and Ellegren 2013). The nature and extent of sex-biased gene expression are quite diverse across taxa and can vary with tissues and developmental stages (Grath and Parsch 2016), and understanding how such dimorphism in expression evolves and is regulated has been a longstanding goal. Some dimorphic genes and traits are directly regulated by the sex-determination cascade. For instance, in *Drosophila*, the acquisition of sexually dimorphic foreleg sex combs is associated with the expression of the sex-determining transcription factor *dsx* (Barmina et al. 2005). Similarly, within the *Drosophila* brain, specific neurons exhibit sex-specific isoforms of the transcription factor *Fru*, a pivotal regulator of sexual dimorphism responsible for directing male-specific courtship behaviors (Dickson 2008). However, the majority of sex-biased genes are not directly bound by members of the sex-determination pathway, and more complex mechanisms must be at play. The configuration of the chromatin structure, which can be open and associated with transcription, or closed and associated with silencing, and can be shaped by the sex-determination pathway (Tachibana 2016), is another regulatory layer that is becoming increasingly appreciated as a source of dimorphism.

In the brown algae *Ectocarpus,* there is a correlation between differential chromatin signatures and sex-biased genes crucial for sexual differentiation (Gueno et al. 2022). Moreover, in mammalian fetal germ cells, the sex-specific chromatin configuration corresponds to anticipated differential gene expression and morphological differences between the sexes during later stages of development (Lesch and Page 2013). Finally, recent research on 2 closely related *Drosophila* species highlighted the involvement of the open chromatin mark H3K4me3 and closed chromatin mark H3K27me2me3 in maintaining sex-biased gene expression (Nanni

et al. 2023). In particular, this work supported the model of "Open in Same sex and/or Closed in Opposite" for sex bias (Nanni et al. 2023), the idea that a gene will be in open chromatin only in the sex where it is primarily expressed (Brown and Bachtrog 2014). However, only 3 histone marks were sampled, and only the head, an organ with limited sexual dimorphism, was considered, such that the relevance of this model to broader patterns of molecular dimorphism is still unclear.

Differences in the male and female chromatin landscape are also crucial to maintain gene expression balance in species with differentiated sex chromosomes, such as the X and Y of mammals. Sex chromosomes originally evolve from ordinary pairs of autosomes after they acquire sex-determining gene(s), which is often followed by recombination suppression over part of the chromosome and degeneration of the Y (in XY system) or W (in ZW systems, where females are ZW while male are ZZ) (D. Charlesworth, Charlesworth, and Marais 2005). The degeneration of the Y (or W) results in different copy numbers of X/Z-linked and autosomal genes in the heterogametic sex, which can cause imbalances in gene expression between sex chromosomes and autosomes (Grath and Parsch 2016). This is thought to be deleterious, as many gene networks and protein complexes involve both X-linked and autosomal gene products (B. Charlesworth 1978; Veitia 2005). Dosage compensation mechanisms have consequently evolved independently in many eukaryotic taxa. The process is thought to begin with the restoration of ancestral expression through increased gene expression on the X (or Z) chromosome in the heterogametic sex. When this increase is not sex-specific, it causes the overexpression of X/Z-linked genes in the homogametic sex, triggering the secondary evolution of mechanisms to suppress X (or Z) expression and restore balance between the sex chromosome and autosomes in the homogametic sex (and consequently between the sexes) (Lucchesi, Kelly, and Panning 2005). Such mechanisms may affect the entire sex chromosome (chromosome-wide dosage compensation) or only a subset of dosage-sensitive genes (gene-by-gene compensation) (Gu and Walters 2017).

Chromosome-wide dosage compensation can therefore be achieved by upregulating X-linked gene expression in the heterogametic sex (e.g. in *Drosophila*) or reducing transcription in the homogametic sex (Muyle et al. 2021). In placental mammals, one of the entire X chromosomes is randomly inactivated in females, while concurrently upregulating a subset of X-linked genes in both sexes (Lucchesi, Kelly, and Panning 2005; O'Neill et al. 2008). In *Caenorhabditis elegans,* downregulation is achieved by a two-fold reduction in transcription rates from both X chromosomes of the homogametic sex (Lau and Csankovszki 2015). Although these compensation mechanisms vary by species, they all lead to the equalization of expression between the sexes, and they are usually accompanied by changes in the sex-specific chromatin structure (Lucchesi and Kuroda 2015; Marin et al. 2017). Studies on dosage compensation have historically been concentrated on species with male heterogamety, which typically show chromosome-wide dosage compensation. On the other hand, ZW systems often have partial dosage compensation mechanisms, where only dosage-sensitive genes are compensated (Itoh et al. 2007; Vicoso et al. 2013). Lepidopteran insects are an important exception to this pattern, as they have a nematode-like dosage compensation that affects the Z chromosome globally (Gu and Walters 2017). However, a recent study on monarch butterflies with young sex chromosomes showed that the ancestral-Z and the neo-Z chromosome unexpectedly showed two distinct modes of dosage compensation mechanisms (Gu et al. 2019). The newly derived Z segment had a *Drosophila*-like dosage compensation mechanism while the transcription of ancestral-Z genes was downregulated to nearly halve the ZZ genes in males as reported in other Lepidopteran species (Gu et al. 2019). Why X/Z:autosome imbalances lead to such diverse

compensation mechanisms is poorly understood, in part due to the limited number of systems that have been characterized in detail.

Here we address these questions in *Artemia franciscana*, a crustacean which exhibits extensive morphological sexual dimorphism (in particular male heads possess protective claspers for mating which are absent in females). A significant number of genes show differential expression in the heads and gonads of males and females, predominantly in gonadal tissue (Huylmans et al. 2019). To understand the molecular basis of this dimorphism, and of crustacean gene regulation in general, we characterized the chromatin landscape of this species using various active and repressive histone marks, in heads and gonads of the two sexes. Furthermore, *Artemia franciscana* has a pair of ZW sex chromosomes which have acquired complete dosage compensation (Huylmans et al. 2019; Bett et al. 2024). However, the underlying molecular mechanisms are unknown. We investigated how epigenetic modifications may mediate equalization of gene expression between sexes despite their differences in gene copy numbers.

Results

Correlation of Histone Modifications with Expression

We generated CUT&TAG datasets from the heads and gonads of *A. franciscana*, focusing on 7 distinct histone modifications. These post-translational histone modifications offer a comprehensive biological annotation across *A. franciscana* genome, encompassing heterochromatin, active transcription, and polycomb-mediated repression (Table 1).

Histone mark	**Active or Repressive**(Kimura 2013)	**No. of replicates in each tissue**	**Known Function** (Bannister and Kouzarides 2011)	**Regulation of sex chromosomes**
H3K27ac	Active	2	Associated with active enhancers	Involved in regulation of sex-determination and differentiation genes in chicken (Jiang et al. 2022)
H3K4me3	Active	2	Marks active promoters	Enriched in the Z-specific region of Schistosome (Picard et al. 2019)
H4K16ac	Active	2	Highly enriched on TSS	Enriched on single male X of Drosophila (Gelbart et al. 2009), neo-Z female of Monarch Moth (Gu et al. 2019)
H3K36me3	Active	2	Enriched toward the 3' of active genes	Transcriptional upregulation of the male X of Drosophila (Bell et al. 2008)
H3K27me3	Repressive	3	Transcriptional repression of developmentally regulated genes	Accumulates on the inactivated X-chromosome of females of placental mammals (Kelsey et al. 2015)
H3K9me2	Repressive	2	Marks constitutive heterochromatin	Enriched on neo-Y regions with a high repeat content in *Drosophila miranda* (Zhou et al. 2013)
H3K9me3	Repressive	2	Associated with heterochromatin, satellite repeats, gene-poor regions	Modulates heterochromatin integrity of Y chromosome (Brown, Nguyen, and Bachtrog 2020)

Table 1: Histone marks profiled in this study together with their role in transcription and associated functions in the regulation of sex chromosomes

Using Spectacle (Song and Chen 2015), we applied a multivariate hidden Markov algorithm to infer 12 chromatin states across the *A. franciscana* genome based on the presence and absence of these 7 histone modifications (supplementary figs. S1, S2, and S3). While these chromatin states are informative about functional elements along the genome, they are difficult to correlate directly with expression, as multiple states can be simultaneously found on each gene. Following Gueno et al. (Gueno et al. 2022), we therefore obtained gene-level chromatin signatures from the combination of states found on each gene. First, the chromatin states (E1 to E12) were clustered into 4 functional categories: (a) active transcription, marked by the enrichment of these 4 active histone modifications or their combinations (H3K36me3, H3K27ac, H3K4me3, and H4K16ac); (b) polycomb-mediated repression, identified by the enrichment of H3K27me3; (c) heterochromatin (if they were enriched for H3K9me2 and/or H3K9me3); (d) mixed chromatin, where both active and repressive chromatin marks are enriched; and we also included a background null state, characterized by the absence of distinct chromatin features. All possible combinations of functional categories of chromatin states, along with the background null state, resulted in a total of 16 gene-level chromatin signatures (Fig. 1a, S1 to S16).

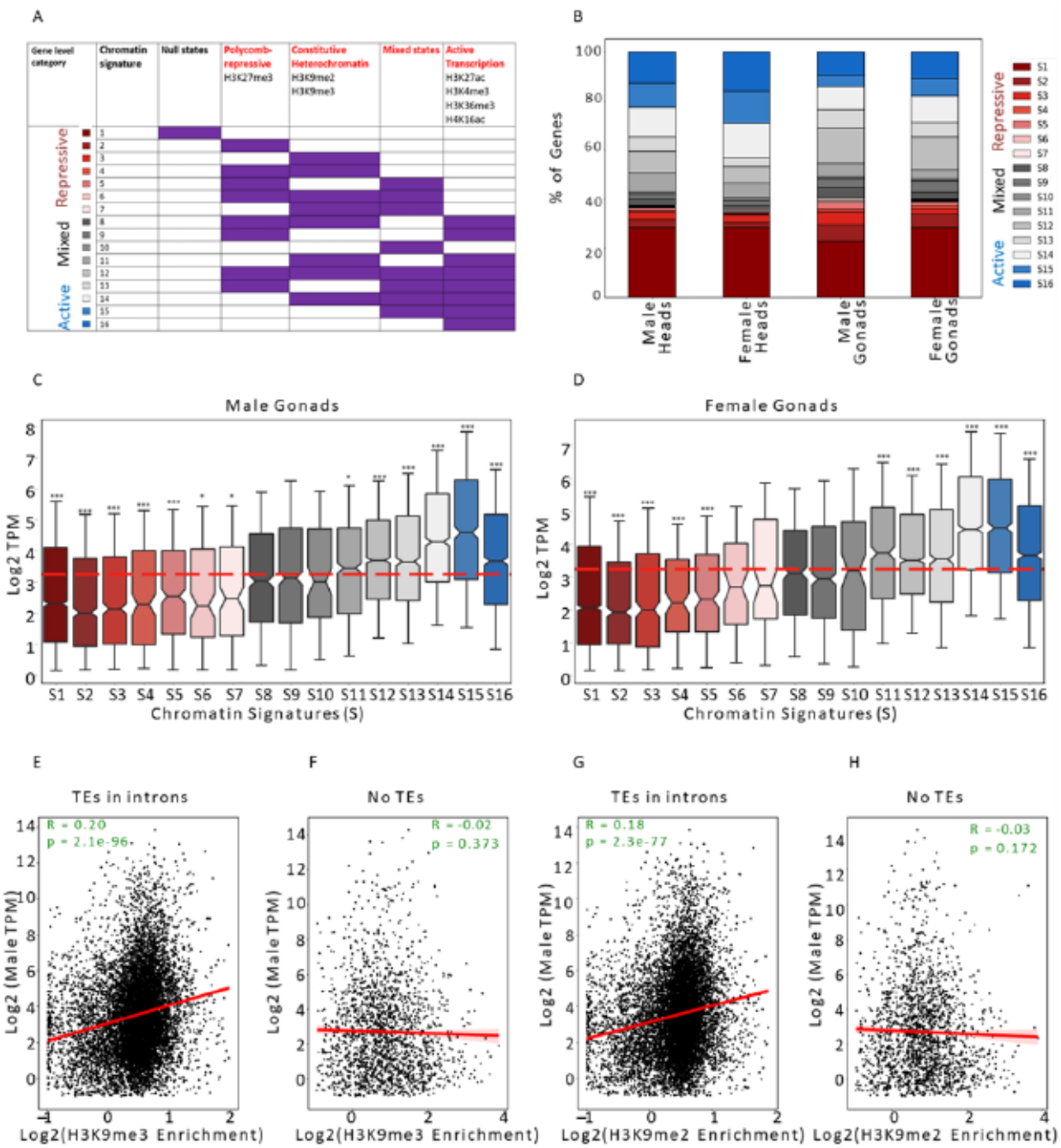


Fig. 1: The chromatin landscape and correlation with expression in *A. franciscana*. a) Active, repressive, and mixed chromatin signatures (S1 to S16) were inferred based on the combination of emission states (supplementary fig. S2) of genes. b) The proportion of chromatin signatures (S1 to S16) across autosomal and pseudoautosomal genes annotated in the *A. franciscana* genome in each tissue and sex. c) Expression of genes with different chromatin signatures (S1 to S16) in male gonads d) Expression patterns of genes of different chromatin signatures (S1 to S16) in female gonads. The dotted line indicates the median expression of all annotated genes. Wilcoxon signed-rank test was performed between genes of particular chromatin signature and the expression of the genes of the rest of the genome (autosomes and pseudoautosomal genes), with * denoting a *P*-value < 0.05, ** is *P*-value < 0.005, and *** is *P*-value < 0.0005. Similar plots are provided for expression in heads in supplementary Figs. 4a and 4b. Panels E and F show the correlation of male head expression with H3K9me3 for genes with e) or without f) TEs in their introns. Panels G and H show the correlation of male head

expression with H3K9me2 for genes with g) or without h) TEs in their introns. The few genes with TEs overlapping their exons were excluded from the analysis. Similar plots using female heads and gonads expression values are provided in supplementary figs. 4C-4N .

These chromatin signatures can then again be broadly classified as active (if the genes contain only active chromatin states and mixed chromatin states, S15 to S16), or repressive (if genes exhibit either the background null state or a combination of repressive and mixed states, S1-S7). Genes with only mixed states, or with both active and repressive marks, were classified as having a mixed chromatin signature (S8 to S14). In order to avoid confounding effects of the sex chromosome chromatin (see section 3), we focused our analysis on genes located in the autosomal and pseudoautosomal regions. Heads were enriched for active chromatin signatures compared to gonads (Fig. 1b, Fisher's Exact, *P*-value = 4.65e-77 in male and *P*-value = 1.31e-142 in female), while repression-associated chromatin signatures were more prevalent in gonads (Fig. 1b, Fisher's exact, *P*-value = 4.81e-07 in male and *P*-value = 4.58e-46 in female). The gonads show enrichment of chromatin signature S12 compared to head tissue. This unusual configuration is linked to genes enriched for protein-binding functions supplementary figs. S21.

In order to explore the functional relevance of this classification, we compared the expression of genes associated with different chromatin signatures in both somatic and gonadal tissues of male and female *A. franciscana*. Genes with repressive signatures (S1 to S7) display significantly lower expression levels (*P*-value < 0.001, Wilcoxon rank sum tests) than genes with other chromatin signatures. In contrast, genes associated with active chromatin signatures (S15 and S16, enriched for H3K36me3, H3K27ac, H3K4me3, and/or H4K16ac) show higher transcript abundance (*P*-value < 0.0001, Wilcoxon rank sum tests; Fig. 1c and 1d, supplementary figs. S4 a and S4b).

Although both H3K27me3 and H3K9me2/me3 are typically repressive (Kimura 2013), H3K9me3 is required for the transcription of certain genes located in heterochromatic regions of the Drosophila genome (Riddle et al. 2011; Ninova, Fejes Tóth, and Aravin 2019; Ninova et al. 2020). This is thought to be driven by the need to repress disruptive transcription from promoters of transposable elements located in their intronic regions (Ninova et al. 2020). The high expression of genes in certain mixed states led us to wonder whether heterochromatin marks might behave similarly in the context of the highly repetitive genome of *A. franciscana* (Bett et al. 2024). In particular, genes in signature S11 (with chromatin marks associated with heterochromatin and any/all active histone marks) and S14 (heterochromatin + active + mixed states) have higher expression than the genome-wide median (*P*-value < 2.6e-63 and *P*-value < 2.8e-115, male and female gonads, respectively, Fig. 1c and 1d, supplementary figs. S4a and S4b), suggesting a relatively permissive role of constitutive heterochromatin in *Artemia*. To investigate if this is related to the presence of transposable element sequences in intronic regions, we separated genes into those carrying at least one transposable element in their introns ("TE in introns"), and those not overlapping with any transposable elements ("No TE"). We then correlated the expression of these two categories of genes with H3K9me3 and H3K9me2 enrichment (Fig. 1 e-f, Fig. 1 g-h, respectively). "TE in introns" genes showed a positive correlation for both histone marks (*P*-value < 8.7e-55 and *P*-value < 2.3e-77, respectively, Spearman correlation), while "No TE" did not. This difference in the correlation coefficients of H3K9me2/3 enrichment and gene expression was significantly different between TE-carrying genes and genes without transposable elements (TEs) for both tissues and sexes (*P*-value < e-10 to e-30 in all cases, Fisher r-to-z transformation of the correlation coefficients, supplementary table S1).

Chromatin Landscape and Sex-biased Gene Expression

We examined how chromatin influences sex-specific gene expression patterns by analyzing two replicated RNA samples from the heads and gonads of both sexes in *A. franciscana*. We again focused only on (pseudo)autosomal genes. Only 10 genes exhibited female bias and 10 genes male bias in heads (adjusted *P*-value < 0.05, fold change > 2, TPM > 0.1; supplementary table S2). No further analyses of sex bias were performed in this tissue, as these small numbers do not allow for meaningful comparisons (but qualitative patterns are similar to gonads, supplementary table S2, , supplementary fig. S5). Gonads, on the other hand, displayed 979 male-biased and 633 female-biased genes (adjusted *P*-value < 0.05, fold change > 2, TPM > 0.1; supplementary table S2). Male-biased genes were more often associated with repressive chromatin signatures (S1 to S7) in ovaries (59%) than in testes (47%, *P*-value < 7.52e-07, Fisher's exact test, Fig. 2a), while the opposite was true for female-biased genes (44% in testis vs. 36% in ovaries, *P*-value < 0.02, Fisher's Exact test, Fig. 2a).

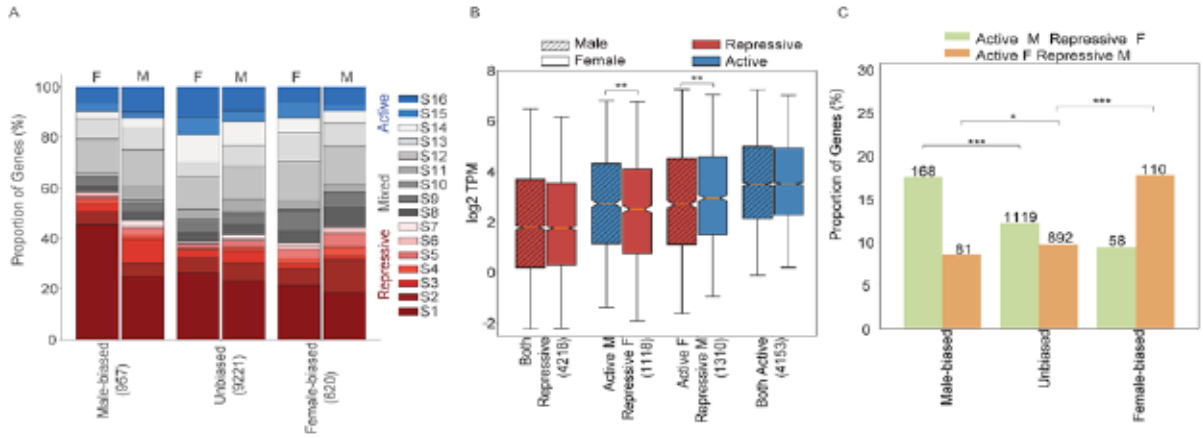


Fig. 2: Sex-biased expression is associated with differential chromatin states in males and females. a) Composition of different categories of sex-biased gene expression in different chromatin signatures in gonads (adjusted *P*-value < 0.05, foldchange >2, and TPM >0.1). b) Expression levels of genes with active-associated or repressive-associated chromatin signatures in both males and females or contrasting chromatin states between the sexes. Statistical significance was evaluated using a pairwise Wilcoxon rank sum test with * denoting a *P*-value < 0.05, ** as *P*-value < 0.005, and *** as *P*-value < 0.0005. c) Proportion of genes with contrasting chromatin states in males and females, for different categories of sex-biased expression in gonads of *A. franciscana* (adjusted *P*-value < 0.05, foldchange >2, and TPM >0.1). Statistical significance was evaluated using a pairwise Wilcoxon rank sum test with * denoting a *P*-value < 0.05, ** as *P*-value < 0.005, and *** as *P*-value < 0.0005.

To directly test whether differences between the male and female chromatin state of each gene influenced their sex-specific pattern of expression, genes were classified based on the active or repressive status of their chromatin signature in the two sexes. In order to be able to classify all sex-biased genes, and because some H3K9me2/3-associated chromatin states were effectively active rather than mixed (see previous section), we classified all chromatin signatures whose associated genes had expression significantly above the genomic median as active (S11 to S16), whereas all others were classified as repressive (Fig. 2b). The "open in same sex/closed in opposite" model (Nanni et al. 2023) predicts that many male-biased genes should show active chromatin signatures in males and repressive signatures in females (and conversely for female-biased genes). This was indeed the case, with female-biased genes for female-active/male-

inactive states (*P*-value < 0.0001 in gonads, chi-square test, Fig. 2c, supplementary figs. S6a-S6b) and male-biased genes being enriched for male-active/female-inactive states (*P*-value < 6.7e-06 in gonads, chi-square test, Fig. 2c, supplementary figs. S6c and S6d). Gene Ontology (GO) enrichment terms of genes that were both female-biased and male-biased with contrasting chromatin-associated signatures between sexes in gonads were enriched in functions related to regulation of gene expression, cellular processes, or system development (supplementary table S3 in female, supplementary table S4 in males).

Epigenetics of Dosage Compensation in *Artemia franciscana*

Artemia franciscana possesses a pair of ZW sex chromosomes, where a well-differentiated region of the Z chromosome (the "S0" region, Fig. 3a) has evolved complete dosage compensation (Huylmans et al. 2019; Bett et al. 2024). In order to determine how chromatin structural changes may regulate dosage compensation in this species, we first assessed the enrichment patterns of normalized histone marks for both replicates across each gene, comparing Z-specific genes (located in the S0 region of the Z chromosome, 139 genes, TPM > 0.5) with autosomal genes in each sex and tissue (supplementary figs. S7 and S8). The normalized signal distribution of most active histone marks (H3K36me3, H3K27ac, and H3K4me3) and repressive histone marks (H3K9me2, H3K9me3, and H3K27me3) were not statistically significantly different between autosomes and the Z-specific region in gonads in both males and females (*P*-value > 0.05, Wilcoxon rank sum tests, supplementary figs. S7 and S8). In heads, only histone marks H3K27ac and H3K9me3 were statistically depleted on the Z-specific region of females compared to the autosomes (*P*-value = 0.028 and *P*-value = 0.002, respectively). However, individual assessments of the two replicates in this female somatic tissue of these histone marks (H3K27ac and H3K9me3) gave inconsistent differences between the autosomes and the Z-specific region, with one of the replicates being significant while the other is insignificant (*P*-value = 0.58 and *P*-value = 1, Wilcoxon rank sum test in H3K27ac and H3K9me3, respectively, supplementary fig. S9), suggesting these may be an artifact rather than true biological signals.

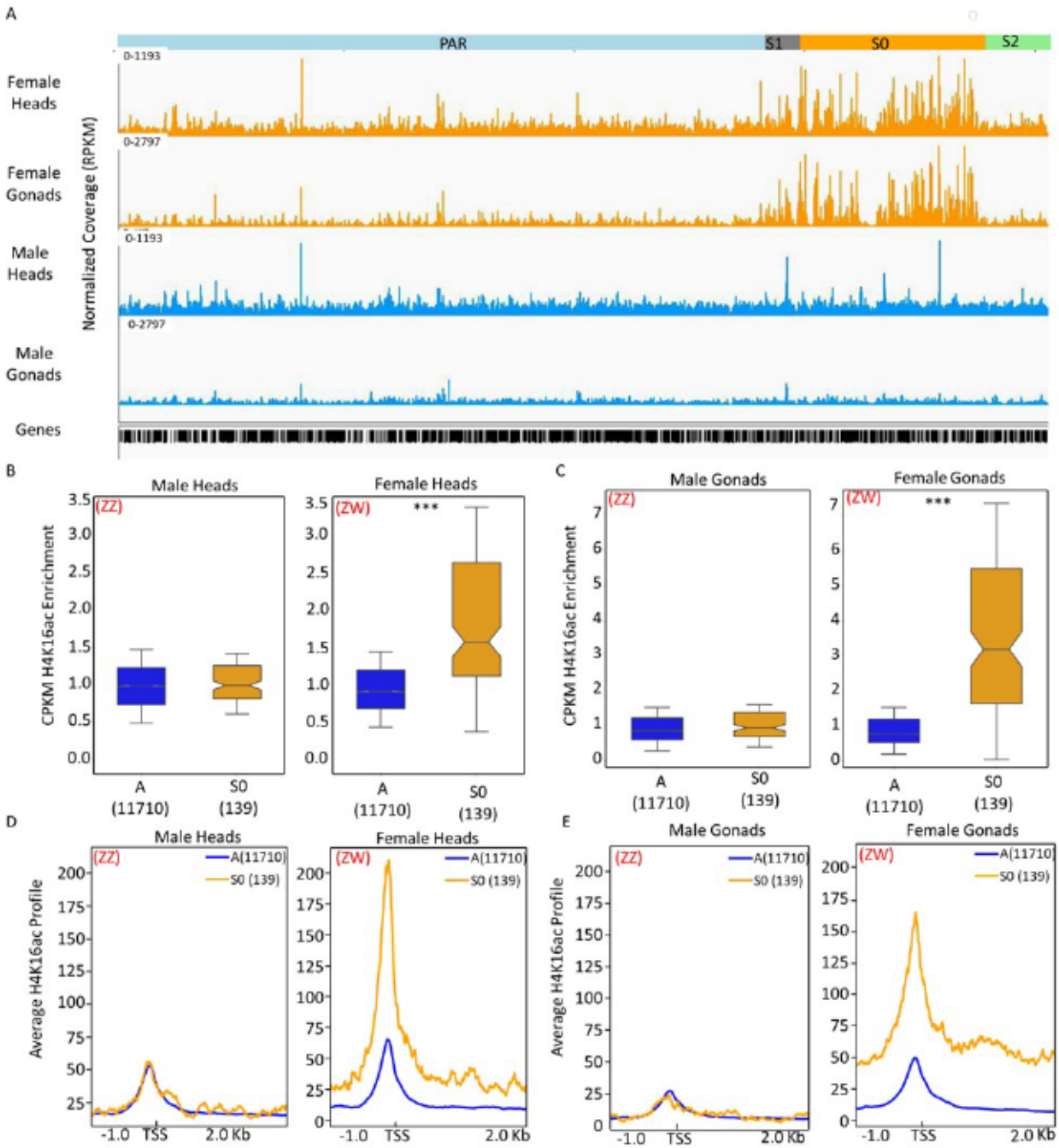


Fig. 3: H4K16ac underlies dosage compensation of the Z. a) Schematic regions of the Z chromosome and genome browser view of H4K16ac in heads and gonads of both males and females. b and c) Boxplots of Normalized H4K16ac enrichment across genes in Z-specific (S0) and autosomes in head b) and gonadal tissue (c). Statistical significance was assessed using a pairwise Wilcoxon rank sum test with * denoting a *P*-value < 0.05, ** as *P*-value < 0.005 and *** as *P*-value < 0.0005. d and e) Average profile enrichment of H4K16ac around TSS (1 kb upstream and 2 kb downstream of TSS) of genes in the Z-specific region (S0) and autosomes of head d) and gonad e) tissues in both males and females.

On the other hand, the active histone mark H4K16ac was enriched approximately two-fold on Z-specific genes (S0) relative to autosomal genes in the heads, and more than two-fold in the gonads in ZW females (*P*-value < 2.2e-24, Wilcoxon rank sum tests, Fig. 3b, Fig. 3c,

supplementary fig. S9). By contrast, the profile of this histone mark was similar between genes in Z-specific and autosomal regions of ZZ males (*P*-value = 1, Wilcoxon rank sum tests, Fig. 3b, Fig. 3c, supplementary fig. S9). To explore how the active chromatin mark H4K16ac is distributed along the length of gene bodies, we examined its average enrichment patterns, spanning 1 kb upstream to 2 kb downstream of the transcription start site (TSS) for genes on both the Z-specific region (S0) and on the autosomes. We used normalized coverage values from replicates from each tissue in each sex. We found that the enrichment is present not only at the TSS but across the entire length of the gene body (Fig. 3d, Fig. 3e). This enrichment of H4K16ac in female relative to male tissues was exclusive to the region (S0) of the Z chromosome (Fig. 3a, supplementary figs. S10 and S11). It should be noted that the role of H4K16ac in *Artemia* dosage compensation has been independently inferred by the Keller-Valsecchi lab (Zimmer et al. 2025).

These patterns may reflect a global enrichment in H4K16ac throughout the S0, or the targeting of a subset of genes by the dosage compensation machinery. To obtain further insights into the architecture of dosage compensation of this group, we first called H4K16ac peaks and found that over 50% of expressed S0 genes overlapped with an H4K16ac peak (84 out of 139 genes, Fig. 4a). Genes overlapping H4K16ac peaks showed increased transcript abundance compared to those outside of peaks (*P*-value = 0.002, Wilcoxon rank sum tests, Fig. 4b), and a trend toward higher sequence conservation (inferred from the nonsynonymous to synonymous divergence ratio, Ka/Ks; the difference was however not significant (Fig. 4d)). However, the expression ratio between females and males did not differ between genes with and without H4K16ac peaks (*P*-value = 0.709, Wilcoxon rank sum tests, Fig. 4c), suggesting that balancing of expression occurs globally for the region.

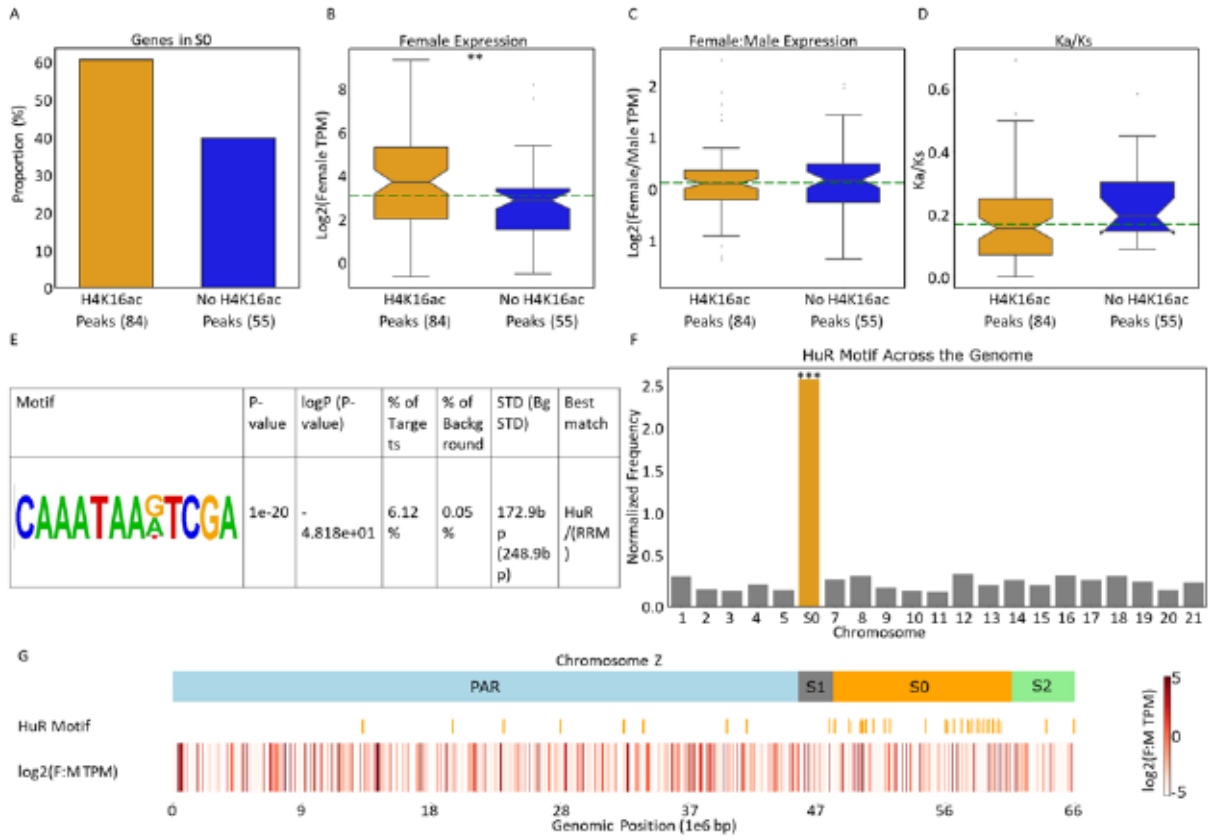


Fig. 4: H4K16ac peaks in the S0 region and their association with functional genomic features. a) Proportion of genes in the Z-specific region (S0) that either intersect or do not intersect with

H4K16ac peaks in female heads. b) Expression of genes in the Z-specific region (S0) with and without H4K16ac peaks in female heads. The significance of the Wilcoxon rank sum test ** is *P*-value < 0.005. c) Log2 of female:male expression ratio of S0 genes with and without H4K16ac peaks in heads. d) The distribution of Ka/Ks of S0 genes with and without overlap with H4K16ac peaks. e) Motif with the most significant enrichment in S0 H4K16ac peaks relative to the rest of the genome. f) Frequency (in number per million base pairs) of the motif described in e) on different chromosomes. *** denotes $P < 0.005$, obtained by resampling N chromosomal loci randomly 10,000 times, where N is the number of peaks found throughout the genome. g) Schematic regions of the Z chromosome, HuR motif distribution patterns, and their relationship to female-to-male expression ratios in head tissue. Similar plots for gonads are provided in supplementary fig. S12.

In order to investigate how H4K16ac may be regulated, we conducted a DNA-binding motif enrichment analysis on the set of H4K16ac peaks (333 and 294 peaks in heads and gonads, respectively) that were intersecting with S0 genes expressed in female tissues (TPM > 0.5 in either female gonads or heads). The HOMER analysis revealed many sequence motifs that were significantly enriched (*P*-value <1e-10) in sequences containing H4K16ac enriched peaks in the Z-specific region (S0) (Fig. 4e, supplementary tables S5 and S6). The most significantly enriched motif in both head and gonad peaks was an AT-rich motif previously identified as a binding motif for Human antigen R (HuR(RRM)) (*P*-value = 1e-19). This motif comprises over 6% of peak regions, but less than 1% of the rest of the genome. Importantly, it is also highly enriched in the S0 region compared to other chromosomes (*P*-value < 0.0001, resampling with 10000 bootstraps, Fig. 4f) and to the rest of the Z chromosome (Fig. 4g), unlike the other 2 most enriched motifs in the peaks. Taken together, these results suggest that a dosage compensation mechanism likely targets the S0 region at specific binding motifs, where it promotes H4K16ac, and global upregulation of gene expression in the region, mirroring the well-known mechanism of Drosophila.

Discussion

Here, we present the first comprehensive analysis of the chromatin landscape of a crustacean species with genetic sex-determination systems. In agreement with the idea that the chromatin landscape may play a regulatory role in gene expression, repressive-associated chromatin signatures were consistently associated with lower transcript abundance compared to active-associated chromatin signatures. Additionally, distinct chromatin states between males and females were associated with sex-biased gene expression in both somatic and gonadal tissues. Finally, we provide evidence that changes in chromatin landscape are likely used in *A. franciscana* to equalize the expression of genes between Z-specific and autosomes and between sexes.

Expression of Gene Regulation in a Highly Repetitive Genome

We sampled a broad range of known active and repressive histone marks in order to get a full overview of a crustacean chromatin landscape. We combined these with published RNA-sequencing data (Huylmans et al. 2019) to characterize epigenetic regulation in this clade. The hatching and growth conditions were slightly different in the 2 studies (in particular, the salinity was 30 to 60 g/l before, and 50 g/l here), but most histone marks yielded the expected associations with expression (supplementary figs. S13 to S17), suggesting this did not strongly impact the conclusions. Interestingly, both repressive marks H3K9me2 and H3K9me3 were overall enriched on genes with high expression (supplementary figs. S18 and S19). Our combinatorial analysis suggests that these two marks of heterochromatin may have distinct

associations with transcription when they act alone (repressive role) or in combination with chromatin marks associated with active transcription (permissive role). This is in line with Feng et al. (Feng et al. 2020) who reported that the genes exhibiting the acquisition of H3K9me3 became heterochromatic, but required the concurrent depletion of H3K4me3 and H3K27ac to undergo transcriptional repression. This combinatorial mode of action may be particularly important in *A. franciscana,* since nearly two-thirds of its genome consist of repetitive elements that need to be condensed while allowing for gene transcription to occur (Bett et al. 2024). Actively transcribed genes situated within constitutive heterochromatin in *Drosophila melanogaster* necessitate this chromatin environment for their appropriate expression, as their activity is compromised upon relocation to open chromatin regions (Ninova, Fejes Tóth, and Aravin 2019). Moreover, the depletion of H3K9me3 leads to the transcriptional suppression of a considerable amount of genes housed within constitutive heterochromatin (Ninova, Fejes Tóth, and Aravin 2019). This may be due to the presence of transposable elements in the introns of these genes, whose promoters need to be repressed through H3K9me3 deposition to avoid spurious intronic transcription (and consequent downregulation of expression of the gene) (Ninova et al. 2020). The presence of TEs within introns in *Artemia franciscana* may play a similarly crucial role in driving the positive correlation between H3K9me2/me3 enrichment and gene expression, supporting the idea that heterochromatization may allow for proper gene regulation in highly repetitive regions. Interestingly, in yeast, H3K9me2 and H3K9me3 are themselves differentially correlated with expression, with H3K9me2 being permissive to transcription and H3K9me3 repressive (Jih et al. 2017). Taken together, these results point to the importance of combinations of histone marks to regulation (i.e. the presence of a "histone code"), but also highlight the evolvability of this code depending on clade and genomic context. Future work on various organisms is required to elucidate how different combinations of histone marks shape gene expression, and the evolutionary and mechanistic pressures driving divergent patterns.

Direct Evidence for the "Open in Same-Sex/Closed in Opposite" Model

If chromatin is associated with sex-specific regulation, sex-biased genes should show sex-specific chromatin marks. In particular, an excess of male-biased genes should be in open chromatin in males but in closed chromatin in females, and vice-versa (Brown and Bachtrog 2014). Brown and Bachtrog (2014) tested this directly in *Drosophila* by comparing the proportion of such sex-inconsistent genes among sex-biased and unbiased genes and found no evidence supporting this hypothesis (Brown and Bachtrog 2014). Recently, Nanni et al. (2023) found that male-biased genes in *Drosophila* heads were enriched in male-open chromatin, and depleted from female-open chromatin (while female-biased genes were enriched in female-open chromatin) (Nanni et al. 2023). However, they did not directly test for a reversal of chromatin state of the same gene between the sexes and were generally limited in the characterization of open and close chromatin by the relatively small number of marks used (one active, H3K4me3, and two repressive, H3K27me2/H3K27me3). Here, we were able to classify our genes as having active or repressed chromatin using 7 different histone marks, and in both heads and gonads of males and females. Very few genes were sex-biased in heads, unlike in the *D. melanogaster* study (Nanni et al. 2023). This is likely due to methodological differences, as Nanni et al. (2023) inferred sex-biased expression at the level of individual exonic features, whereas, in *Artemia*, sex-biased genes were identified based on the cumulative expression levels across all exons of a gene, and only if they had at least a two-fold change, a much more conservative approach. This precluded the detection of very subtle changes between the sexes, which typically comprise the bulk of sex-biased genes in *Drosophila* head and brain datasets (Huylmans and Parsch 2015). Whether such small changes are biologically relevant is not clear,

and we therefore focused on more robust sex-biased expression, which is largely found in gonads. Our detailed classification showed that many genes have sex-inconsistent epigenetic states in the gonad, and these are enriched among sex-biased genes in the predicted direction. This provides a more direct test of the model than previously possible and strongly supports a contribution of the chromatin landscape to sex-specific regulation of expression.

Artemia franciscana Exhibits *Drosophila*-like Dosage Compensation

Early large-scale gene expression studies highlighted an apparent discrepancy between ZW and XY species, with only the latter showing dosage compensation at the chromosome level (Gu and Walters 2017). Since then, multiple studies in Lepidoptera have shown that dosage compensation can regulate whole Z chromosomes, but this has so far remained exceptional among ZW systems (Walters, Hardcastle, and Jiggins 2015; Gu, Walters, and Knipple 2017). *A. franciscana* shows a significant enrichment of H4K16ac on the Z-specific regions of the Z chromosome compared to autosomes in females across both head and gonad tissues, consistent with dosage compensation acting to increase gene expression of the female Z. It should be noted that because of the small number of genes in the S0 (~1% of all genes), subtler changes in other chromatin marks may not be as easily detected and cannot be fully excluded. The female enrichment in H4K16ac is in line with earlier results on gene expression, which showed that Z-specific and autosomal genes have similar expression levels in both males and females (Huylmans et al. 2019; Bett et al. 2024). This mechanism resembles the mode of action of dosage compensation in *Drosophila*, where the dosage compensation complex deposits active H4K16ac histone marks on the single male X-chromosome (Lucchesi and Kuroda 2015). Another interesting parallel is the presence of a candidate motif that is enriched in the differentiated region of the Z and may be targeted by the dosage compensation machinery, as suggested by its frequent co-occurrence with H4K16ac peaks. This is reminiscent of the high-affinity sites of *D. melanogaster*, which are enriched for a guanine-adenine (GA)-rich motif that allows for their recognition and binding by the dosage compensation complex (Alekseyenko et al. 2008; Straub et al. 2008). H4K16ac-mediated dosage compensation has also been observed in green anole lizards, where males exhibit an enrichment of H4K16ac on their X-chromosome (Marin et al. 2017), although in this case recognition and binding to the X-chromosome is mediated by a long non-coding RNA (Tenorio et al. 2024). Contrary to the original dichotomy between ZW and XY systems, these results therefore not only suggest that chromosome-wide compensation of Z chromosomes may be more common than previously thought, but also highlight a striking case of molecular convergence of the regulation of the 2 types of sex chromosomes. Future work in diverse ZW and XY systems is needed to shed light on how common different types of compensation are, and what evolutionary scenarios lead to each of them.

Methods

Dissection of Heads and Gonads of *Artemia franciscana*

Artemia franciscana cysts were purchased by the Fish Facility of the Institute of Science and Technology Austria from the Zierfischfutter Wünnenberg (who source them from Sanders), and hatched at 24 to 26 °C, in 550 ml red sea salt in 24L in 14:10 h light/darkness. Freshly hatched nauplii were transferred to the lab and reared until adulthood in individual vials to avoid mating under 50 grams/litres salinity and in 16:08 h light/darkness. Salt was washed off adult *A. franciscana* with deionized water and 1 × PBS (without $MgCl_2$ and $CaCl_2$), and tissues were dissected under cold DPBS. The ovisac contents from the female gonads were emptied and discarded to minimize the effect of the female germline on our results, as germ cells have recently been shown to lack dosage compensation in *A. franciscana* (Elkrewi and Vicoso

2024). Tissues from 4 adults per replicate were then permeabilized for 30 min in an ice-cold mixture of 10 mg of Collagenase and 1 ml of Enzyme Dissociation Buffer.

CUT&Tag Library Preparation

CUT&TAG assays were performed using CUT&Tag-IT™ Assay Kit (Active Motif cat no. 53160). Permeabilized tissues were mixed with activated Concanavalin A beads after their resuspension in the wash buffer. We then incubated overnight at 4 °C with primary antibodies from Active Motif: H3K27me3 (cat no. 39157), H3K36me3 (cat no. 61102), H3K4me3 (cat no. 39160), H3K27ac (cat no. 39034), H4K16ac (cat no. 39930), H3K9me3 (cat no. 39766) and H3K9me2 (cat no. 39754). Following this, samples were incubated with a diluted Guinea Pig Anti-Rabbit secondary antibody for 1 h in a nutator at room temperature. The beads were then washed with a Dig-Wash buffer, and assembled pA-Tn5 transposomes were then added and incubated at room temperature for 1 h in a nutator mixer. Following incubation, washing of the beads was performed using a Dig-300 buffer. Tagmentation was completed by adding a tagmentation buffer that was mixed with Protease Inhibitor Cocktail and 5% Digitonin and then incubating them at 37 °C for 80 min. In order to halt the tagmentation process and digest DNA fragments, 4.2 µl of EDTA, 1.25 µl of 10% SDS, and 1.1 µl of Proteinase K were added to the Tagmentation buffer and incubated for 60 min at 55 °C. DNA was extracted and then amplified in a 50 µl PCR reaction using a combination of unique i7 Indexed primer and one i5 Indexed primer, 30 ul of tagmented DNA, 1 ul of 10 mM dNTPs, 10 ul of 5× Q5 Reaction Buffer, and 0.5 ul of Q5 DNA polymerase. The desired fragment size libraries of CUT&TAG were then selected using 1.1 × of solid-phase reversible immobilization (SPRI) beads to DNA (55µl of SPRI was added to 50 µl of DNA). The libraries were checked for quality using both Agilent Bioanalyzer for fragment size distribution and qPCR for DNA library concentration, and those that passed minimum quality checks (with at least 2.5 nM in concentration) were sequenced on Illumina NextSeq550 PE150 Medium at the Vienna Biocenter sequencing facility.

Processing of CUT&TAG Raw Reads

Paired-end raw reads were checked for quality with fastqc (Andrews 2010). Overrepresented sequences were filtered out using the bbduk.sh package of bbmap tool (Bushnell 2014) with the options "ref = overrepresented.fasta ktrim = r k = 31 mink = 11 hdist = 1 tpe tbo." Sequencing adapters were then trimmed with cutadapt under parameters "-a CTGTCTCTTATACAC -A CTGTCTCTTATACAC -O 5 -q 0 -m 20 -j 30" (Martin 2011). The trimmed paired-end reads were aligned to *A. franciscana* reference genome (Bett et al. 2024) using bowtie2 v2.4.5 (Langmead and Salzberg 2012) with these options: "–local –very-sensitive-local –no-unal –no-mixed –no-discordant –phred33 -I 10 -X 700." The alignments were sorted by coordinates using Picard ("Picard Toolkit" 2019) with the parameter "-SORT_ORDER coordinate." PCR duplicates were then marked and removed with Picard using the options "REMOVE_DUPLICATES = true, ASSUME_SORTED = TRUE." The resulting alignments were filtered further with the options "-q 2 -F 0 × 04 -f 0 × 2" in samtools v1.18 (Li et al. 2009) and used for downstream analysis (supplementary table S7). Consistency between biological replicates was addressed by calculating the pairwise Pearson correlation coefficient of the normalized coverage along TEs and genes, for all of the samples using the R function "cor" from the package stats (supplementary fig. S20).

Expression Analysis

The RNA-seq samples that were obtained from the heads and gonads of 10 adult males and 10 adult females of *A. franciscana* (Huylmans et al. 2019) were mapped (supplementary table S8) using Kallisto (Bray et al. 2016) to the annotated coding sequences (CDs) of *A. franciscana*

(from annotation (Bett et al. 2024)) to create pseudoalignment and were then imported to sleuth (Pimentel et al. 2017) for estimation of transcript abundance. The gene expression within each tissue across sexes was normalized using NormalyzerDE (Willforss, Chawade, and Levander 2019) in Rv4.1.2. We then performed differential expression analysis with the sleuth package using a likelihood ratio test (LRT). Genes that exhibited more than 2-fold change (FC) difference with a false discovery rate (FDR) of <0.05 in expression were categorized as male or female-biased depending on the direction of the change.

Identification of Optimal Chromatin State Number

We used Spectacle to assign chromatin states along the *A. franciscana* genome (Song and Chen 2015). First, alignment of CUT&TAG reads from each of the 7 histone marks was converted to bed format with bedtools and then binarized across the genome in a bin size of 200 bp using BinarizedBed with option "-strictthresh." Second, binarized data were then used in the learning model of each given chromatin state number. This was performed with LearnModel with the options '-i spectral -comb' to allow the hidden Markov model to use a spectral learning algorithm in its chromatin state estimation. Next, we employed the CompareModels function within Spectacle to assess the maximum emission parameters, correlating them with selected chromatin states using Pearson correlation. Specifically, we examined the correlation of chromatin state 21 with a range of simpler models, encompassing chromatin state numbers from 20 to 2. The median of these correlations was plotted to pinpoint the optimal chromatin state number (supplementary fig. S1). We identified 12 chromatin states as the optimal state number, as any further increase in the number of chromatin states only lead to the recovery of states with similar characteristics to the starting chromatin state 21. Therefore, chromatin state number 12 was selected for downstream analysis.

Analysis of Chromatin States

The coverage of each emission state (chromatin state 12) across the *A. franciscana* genome was generated using the MakeSegmentation function. The outputs were then intersected with the coordinates of genes using the bedtools intersect function under default settings. Emission states were further summarized into 3 major groups depending on their functional annotation (Brown and Bachtrog 2014); (i) active transcription (H3K27ac, H4K16ac, H3K4me3, and H3K36me3), (ii) polycomb-mediated repression (H3K27me3) and (iii) constitutive heterochromatin (H3K9me3 and H3K9me2). Additionally, the inclusion of mixed states that contain combination of either facultative or constitutive histone marks (H3K9me3, H3K9me2, and H3K27me3) with any of the active transcription marks (H3K36me3, H3K27ac, H3K16ac, and H3K4me3) and Null state resulted in definition of 16 chromatin signatures as done in (Gueno et al. 2022).

Correlating Genes With Chromatin Marks

To assess the enrichment of chromatin marks across genes, we adopted a similar approach to that of Anderson et al. (2023) by tallying the mapped reads spanning from the beginning to the end of every gene in each replicate. This was accomplished by aggregating reads with deepTools/multiBamSummary using a bed annotation file and then normalizing counts based on the number of mapped reads and the gene length in kilobases in Million. The resulting metric provides counts per kilobase per million (CPKM).

> "*multiBamSummary BED-file –BED annotation.bed –bamfiles alignment.bam –centerReads –maxFragmentLength 700 -o results.npz –outRawCounts readCounts.tab*"

To generate average plots depicting chromatin signals surrounding the TSS of genes of each replicate, we applied deepTools/bamcoverage to summarize normalized coverage within a window of 10-bp intervals "*bamCoverage -b alignment.bam –binSize 10 -p 20 –normalizeUsing RPKM –maxFragmentLength 700 -o file.bw*".

To average the distribution of chromatin marks of surrounding TSS of genes in both replicates, we used bamCompare function of deepTools "*bamCompare –bamfile1 file1.bam –bamfile2 file2.bam -p 20 -bs 5 –normalizeUsing RPKM –skipNAs –minFragmentLength 10 –maxFragmentLength 700 –operation mean –scaleFactorsMethod None -o output.bw.*"

The outputs were used as inputs for computing matrices using deepTools/computeMatrix "*computeMatrix reference-point –referencePoint TSS -S file.bw -R annotation.bed –missingDataAsZero -b 1000 -a 2000 –binSize 5 -o file.mat.gz*" and visualized using plotProfile feature.

In order to visualize the distribution of H4K16ac in each tissue across all 21 chromosomes of *A. franciscana*, we used genome coverage data generated using deepTools/bamCoverage with these options: "*bamCoverage -b file.bam -bs 30000 -p 20 –outFileFormat bedgraph –normalizeUsing RPKM –maxFragmentLength 700 -o file.bedgraph.*"

Peaks Calling and Motif Enrichment

The deduplicated alignments generated earlier were converted to BED format using the bamToBed function from Bedtools (Quinlan and Hall 2010). Read pairs mapping to the same chromosome with fragment lengths under 700 bp were selected for the generation of bedgraph files with the genomecov function in Bedtools (Quinlan and Hall 2010). Peaks were then called using SEACR [Sparse Enrichment Analysis for CUT&RUN v1.3 (Meers, Tenenbaum, and Henikoff 2019)]. The SEACR peaks were called using "SEACR_1.3.sh sample.bedgraph 0.01 non stringent sample_seacr_top0.01.peaks." The H4K16ac peaks corresponding to selected genes (TPM > 0.5) of Z-specific region (S0) in females were used for sequence motifs identification using HOMER (Hypergeometric Optimization of Motif EnRichment (Heinz et al. 2010)). Motifs were identified as follows; "findMotifsGenome.pl H4K16acpeaks.bed Afranciscana_genome.fa.masked S0peaks_homerMotif -mask -size 1000 -p 20." In order to identify instances of the targeted motif of interest, we used scanMotifGenomeWide.pl function of HOMER as follows: scanMotifGenomeWide.pl S0peaks_homerMotif/homerResults/motif1.motif Afranciscana_genome.fa.masked -mask -keepAll -bed > S0peaks_homerMotif/motif1.sites.bed

Estimating Rates of Evolution of Genes With and Without H4K16ac Peaks

The CDS of annotated genes from *Artemia franciscana* (Bett et al. 2024) and their homologous sequences of annotated genes from *Artemia sinica* (Elkrewi et al. 2022) were aligned using the TranslatorX software package (Abascal, Zardoya, and Telford 2010) with the "gblocks" option to filter out poorly aligned or unreliable regions, ensuring a high-quality alignment for downstream analyses. Following the alignment, the nonsynonymous (Ka) and synonymous (Ks) substitution rates were calculated using the KaKs_calculator (Wang et al. 2010), employing the Yang and Nielsen (YN) algorithm.

Acknowledgments

We thank the Vicoso lab for their help in maintaining *Artemia* and for their valuable feedback and suggestions. We thank Marwan Elkrewi for his useful technical advice and discussions. We are also grateful to the Scientific Unit at ISTA Austria for computational resources and assistance. This work was supported by Austrian science fund (FWF) grants PAT8748323 and SFB F88-10 (as part of the SFB Meiosis consortium https://sfbmeiosis.org) to BV and Swedish Research Council (Vetenskapsrådet, grant number 2020-06424) to MSTA.

Data Availability

The CUT&TAG raw reads generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) BioProject under accession number PRJNA1150095. The Bioinformatics pipeline that was used for processing and analysis of CUT&TAG sequencing reads can be accessed on this gitpage (https://github.com/vkb25/Chromatin-landscape-in-Artemia-franciscana.git).

References

Abascal, Federico, Rafael Zardoya, and Maximilian J. Telford. 2010. "TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations." *Nucleic Acids Research* 38 (Web Server issue): W7-13. https://doi.org/10.1093/nar/gkq291.

Alekseyenko, Artyom A., Shouyong Peng, Erica Larschan, Andrey A. Gorchakov, Ok-Kyung Lee, Peter Kharchenko, Sean D. McGrath, et al. 2008. "A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome." *Cell* 134 (4): 599–609. https://doi.org/10.1016/j.cell.2008.06.033.

Anderson, James, Steven Henikoff, and Kami Ahmad. 2023. "Chromosome-Specific Maturation of the Epigenome in the Drosophila Male Germline," February. https://doi.org/10.1101/2023.02.24.529909.

Andrews, S. 2010. "FASTQC. A Quality Control Tool for High Throughput Sequence Data."

Bannister, Andrew J, and Tony Kouzarides. 2011. "Regulation of Chromatin by Histone Modifications." *Cell Research* 21 (3): 381–95. https://doi.org/10.1038/cr.2011.22.

Barmina, Olga, Martin Gonzalo, Lauren M. McIntyre, and Artyom Kopp. 2005. "Sex- and Segment-Specific Modulation of Gene Expression Profiles in Drosophila." *Developmental Biology* 288 (2): 528–44. https://doi.org/10.1016/j.ydbio.2005.09.052.

Bell, Oliver, Thomas Conrad, Jop Kind, Christiane Wirbelauer, Asifa Akhtar, and Dirk Schübeler. 2008. "Transcription-Coupled Methylation of Histone H3 at Lysine 36 Regulates Dosage Compensation by Enhancing Recruitment of the MSL Complex in Drosophila Melanogaster." *Molecular and Cellular Biology* 28 (10): 3401–9. https://doi.org/10.1128/MCB.00006-08.

Bett, Vincent Kiplangat, Ariana Macon, Beatriz Vicoso, and Marwan Elkrewi. 2024. "Chromosome-Level Assembly of Artemia Franciscana Sheds Light on Sex-Chromosome Differentiation." *Genome Biology and Evolution*, January, evae006. https://doi.org/10.1093/gbe/evae006.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34 (5): 525–27. https://doi.org/10.1038/nbt.3519.

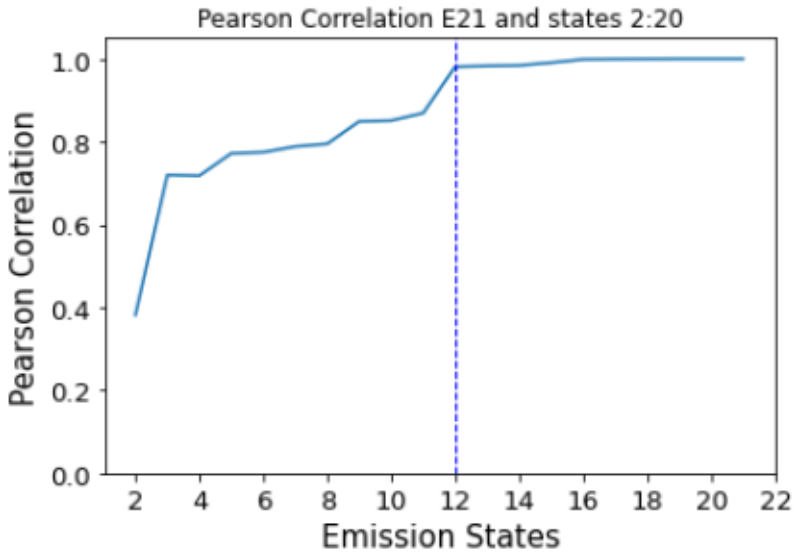
Brown, Emily J., and Doris Bachtrog. 2014. "The Chromatin Landscape of Drosophila: Comparisons between Species, Sexes, and Chromosomes." *Genome Research* 24 (7): 1125–37. https://doi.org/10.1101/gr.172155.114.

Brown, Emily J, Alison H Nguyen, and Doris Bachtrog. 2020. "The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide." *Molecular Biology and Evolution* 37 (10): 2808–24. https://doi.org/10.1093/molbev/msaa082.
Bushnell, Brian. 2014. "BBMap: A Fast, Accurate, Splice-Aware Aligner," March. https://www.osti.gov/biblio/1241166.
Charlesworth, B. 1978. "Model for Evolution of Y Chromosomes and Dosage Compensation." *Proceedings of the National Academy of Sciences of the United States of America* 75 (11): 5618–22. https://doi.org/10.1073/pnas.75.11.5618.
Charlesworth, D., B. Charlesworth, and G. Marais. 2005. "Steps in the Evolution of Heteromorphic Sex Chromosomes." *Heredity* 95 (2): 118–28. https://doi.org/10.1038/sj.hdy.6800697.
Dickson, Barry J. 2008. "Wired for Sex: The Neurobiology of Drosophila Mating Decisions." *Science* 322 (5903): 904–9. https://doi.org/10.1126/science.1159276.
Elkrewi, Marwan, Uladzislava Khauratovich, Melissa A. Toups, Vincent Kiplangat Bett, Andrea Mrnjavac, Ariana Macon, Christelle Fraisse, et al. 2022. "ZW Sex-Chromosome Evolution and Contagious Parthenogenesis in Artemia Brine Shrimp." *Genetics* 222 (2). https://doi.org/10.1093/genetics/iyac123.
Elkrewi, Marwan, and Beatriz Vicoso. 2024. "Single-Nucleus Atlas of the Artemia Female Reproductive System Suggests Germline Repression of the Z Chromosome." *PLOS Genetics* 20 (8): e1011376. https://doi.org/10.1371/journal.pgen.1011376.
Feng, Ying, Yao Wang, Xiangnan Wang, Xiaohui He, Chen Yang, Ardalan Naseri, Thoru Pederson, et al. 2020. "Simultaneous Epigenetic Perturbation and Genome Imaging Reveal Distinct Roles of H3K9me3 in Chromatin Architecture and Transcription." *Genome Biology* 21 (1): 296. https://doi.org/10.1186/s13059-020-02201-1.
Gelbart, Marnie E., Erica Larschan, Shouyong Peng, Peter J. Park, and Mitzi I. Kuroda. 2009. "Drosophila MSL Complex Globally Acetylates H4K16 on the Male X Chromosome for Dosage Compensation." *Nature Structural & Molecular Biology* 16 (8): 825–32. https://doi.org/10.1038/nsmb.1644.
Grath, Sonja, and John Parsch. 2016a. "Sex-Biased Gene Expression." *Annual Review of Genetics* 50 (1): 29–44. https://doi.org/10.1146/annurev-genet-120215-035429.
Gu, Liuqi, Patrick F. Reilly, James J. Lewis, Robert D. Reed, Peter Andolfatto, and James R. Walters. 2019. "Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly." *Current Biology* 29 (23): 4071-4077.e3. https://doi.org/10.1016/j.cub.2019.09.056.
Gu, Liuqi, and James R. Walters. 2017a. "Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details." *Genome Biology and Evolution* 9 (9): 2461–76. https://doi.org/10.1093/gbe/evx154.
Gu, Liuqi, and James R Walters. 2017b. "Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details." *Genome Biology and Evolution* 9 (9): 2461–76. https://doi.org/10.1093/gbe/evx154.
Gu, Liuqi, James R. Walters, and Douglas C. Knipple. 2017. "Conserved Patterns of Sex Chromosome Dosage Compensation in the Lepidoptera (WZ/ZZ): Insights from a Moth Neo-Z Chromosome." *Genome Biology and Evolution* 9 (3): 802–16. https://doi.org/10.1093/gbe/evx039.
Gueno, Josselin, Michael Borg, Simon Bourdareau, Guillaume Cossard, Olivier Godfroy, Agnieszka Lipinska, Leila Tirichine, J. Mark Cock, and Susana M. Coelho. 2022. "Chromatin Landscape Associated with Sexual Differentiation in a UV Sex Determination System." *Nucleic Acids Research* 50 (6): 3307–22. https://doi.org/10.1093/nar/gkac145.

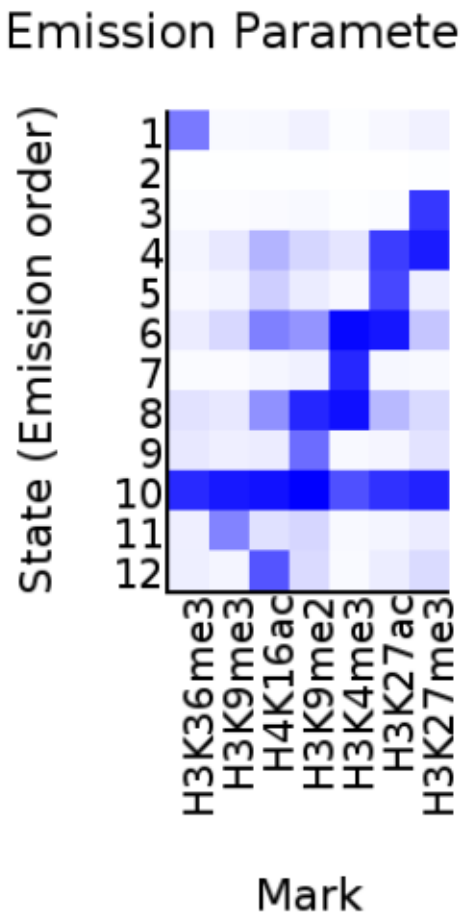
Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities." *Molecular Cell* 38 (4): 576–89. https://doi.org/10.1016/j.molcel.2010.05.004.
Huylmans, Ann Kathrin, and John Parsch. 2015. "Variation in the X:Autosome Distribution of Male-Biased Genes among Drosophila Melanogaster Tissues and Its Relationship with Dosage Compensation." *Genome Biology and Evolution* 7 (7): 1960–71. https://doi.org/10.1093/gbe/evv117.
Huylmans, Ann Kathrin, Melissa A. Toups, Ariana Macon, William J. Gammerdinger, and Beatriz Vicoso. 2019. "Sex-Biased Gene Expression and Dosage Compensation on the Artemia Franciscana Z-Chromosome." *Genome Biology and Evolution* 11 (4): 1033–44. https://doi.org/10.1093/gbe/evz053.
Itoh, Yuichiro, Esther Melamed, Xia Yang, Kathy Kampf, Susanna Wang, Nadir Yehya, Atila Van Nas, et al. 2007. "Dosage Compensation Is Less Effective in Birds than in Mammals." *Journal of Biology* 6 (1): 2. https://doi.org/10.1186/jbiol53.
Jiang, Yunqi, Zhelun Peng, Qiu Man, Sheng Wang, Xiaochen Huang, Lu Meng, Heng Wang, and Guiyu Zhu. 2022. "H3K27ac Chromatin Acetylation and Gene Expression Analysis Reveal Sex- and Situs-Related Differences in Developing Chicken Gonads." *Biology of Sex Differences* 13 (1): 6. https://doi.org/10.1186/s13293-022-00415-5.
Kelsey, Angela D., Christine Yang, Danny Leung, Jakub Minks, Thomas Dixon-McDougall, Sarah E. L. Baldry, Aaron B. Bogutz, Louis Lefebvre, and Carolyn J. Brown. 2015. "Impact of Flanking Chromosomal Sequences on Localization and Silencing by the Human Non-Coding RNA XIST." *Genome Biology* 16 (October):208. https://doi.org/10.1186/s13059-015-0774-2.
Kimura, Hiroshi. 2013. "Histone Modifications for Human Epigenome Analysis." *Journal of Human Genetics* 58 (7): 439–45. https://doi.org/10.1038/jhg.2013.66.
Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.
Lau, Alyssa C., and Györgyi Csankovszki. 2015. "Balancing up and Downregulation of the C. Elegans X Chromosomes." *Current Opinion in Genetics & Development* 31 (April):50–56. https://doi.org/10.1016/j.gde.2015.04.001.
Lesch, Bluma J., and David C. Page. 2013. "Sex-Specific Chromatin States in Mammalian Fetal Germ Cells." *Epigenetics & Chromatin* 6 (1): P45. https://doi.org/10.1186/1756-8935-6-S1-P45.
Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.
Lucchesi, John C., William G. Kelly, and Barbara Panning. 2005. "Chromatin Remodeling in Dosage Compensation." *Annual Review of Genetics* 39:615–51. https://doi.org/10.1146/annurev.genet.39.073003.094210.
Lucchesi, John C., and Mitzi I. Kuroda. 2015. "Dosage Compensation in Drosophila." *Cold Spring Harbor Perspectives in Biology* 7 (5). https://doi.org/10.1101/cshperspect.a019398.
Marin, Ray, Diego Cortez, Francesco Lamanna, Madapura M. Pradeepa, Evgeny Leushkin, Philippe Julien, Angélica Liechti, et al. 2017a. "Convergent Origination of a Drosophila-like Dosage Compensation Mechanism in a Reptile Lineage." *Genome Research* 27 (12): 1974–87. https://doi.org/10.1101/gr.223727.117.
Marin R, Cortez D, Lamanna F, Pradeepa MM, Leushkin E, Julien P, Liechti A, Halbert J, Brüning T, Mössinger K, et al. 2017b. "Convergent Origination of a Drosophila-like Dosage

Compensation Mechanism in a Reptile Lineage." *Genome Research* 27 (12): 1974–87. https://doi.org/10.1101/gr.223727.117.
Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO - 10.14806/Ej.17.1.200*, May. https://journal.embnet.org/index.php/embnetjournal/article/view/200.
Meers, Michael P., Dan Tenenbaum, and Steven Henikoff. 2019. "Peak Calling by Sparse Enrichment Analysis for CUT&RUN Chromatin Profiling." *Epigenetics & Chromatin* 12 (1): 42. https://doi.org/10.1186/s13072-019-0287-4.
Muyle, Aline, Doris Bachtrog, Gabriel A. B. Marais, and James M. A. Turner. 2021. "Epigenetics Drive the Evolution of Sex Chromosomes in Animals and Plants." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1826): 20200124. https://doi.org/10.1098/rstb.2020.0124.
Nanni, Adalena V, Natalie Martinez, Rita Graze, Alison Morse, Jeremy R B Newman, Vaibhav Jain, Srna Vlaho, et al. 2023. "Sex-Biased Expression Is Associated With Chromatin State in Drosophila Melanogaster and Drosophila Simulans." *Molecular Biology and Evolution* 40 (5): msad078. https://doi.org/10.1093/molbev/msad078.
Ninova, Maria, Katalin Fejes Tóth, and Alexei A. Aravin. 2019. "The Control of Gene Expression and Cell Identity by H3K9 Trimethylation." *Development (Cambridge, England)* 146 (19). https://doi.org/10.1242/dev.181180.
Ninova, Maria, Baira Godneeva, Yung-Chia Ariel Chen, Yicheng Luo, Sharan J. Prakash, Ferenc Jankovics, Miklós Erdélyi, Alexei A. Aravin, and Katalin Fejes Tóth. 2020. "The SUMO Ligase Su(Var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation." *Molecular Cell* 77 (3): 571-585.e4. https://doi.org/10.1016/j.molcel.2019.09.033.
O'Neill, Laura P., Hugh T. Spotswood, Milan Fernando, and Bryan M. Turner. 2008. "Differential Loss of Histone H3 Isoforms Mono-, Di- and Tri-Methylated at Lysine 4 during X-Inactivation in Female Embryonic Stem Cells." *Biological Chemistry* 389 (4): 365–70. https://doi.org/10.1515/BC.2008.046.
Parsch, John, and Hans Ellegren. 2013. "The Evolutionary Causes and Consequences of Sex-Biased Gene Expression." *Nature Reviews. Genetics* 14 (2): 83–87. https://doi.org/10.1038/nrg3376.
Picard, Marion A. L., Beatriz Vicoso, David Roquis, Ingo Bulla, Ronaldo C. Augusto, Nathalie Arancibia, Christoph Grunau, Jérôme Boissier, and Céline Cosseau. 2019. "Dosage Compensation throughout the Schistosoma Mansoni Lifecycle: Specific Chromatin Landscape of the Z Chromosome." *Genome Biology and Evolution* 11 (7): 1909–22. https://doi.org/10.1093/gbe/evz133.
"Picard Toolkit." 2019. *Broad Institute, GitHub Repository*. Broad Institute. https://broadinstitute.github.io/picard/.
Pimentel, Harold, Nicolas L. Bray, Suzette Puente, Páll Melsted, and Lior Pachter. 2017. "Differential Analysis of RNA-Seq Incorporating Quantification Uncertainty." *Nature Methods* 14 (7): 687–90. https://doi.org/10.1038/nmeth.4324.
Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics (Oxford, England)* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.
Riddle, Nicole C., Aki Minoda, Peter V. Kharchenko, Artyom A. Alekseyenko, Yuri B. Schwartz, Michael Y. Tolstorukov, Andrey A. Gorchakov, et al. 2011. "Plasticity in Patterns of Histone Modifications and Chromosomal Proteins in Drosophila Heterochromatin." *Genome Research* 21 (2): 147–63. https://doi.org/10.1101/gr.110098.110.

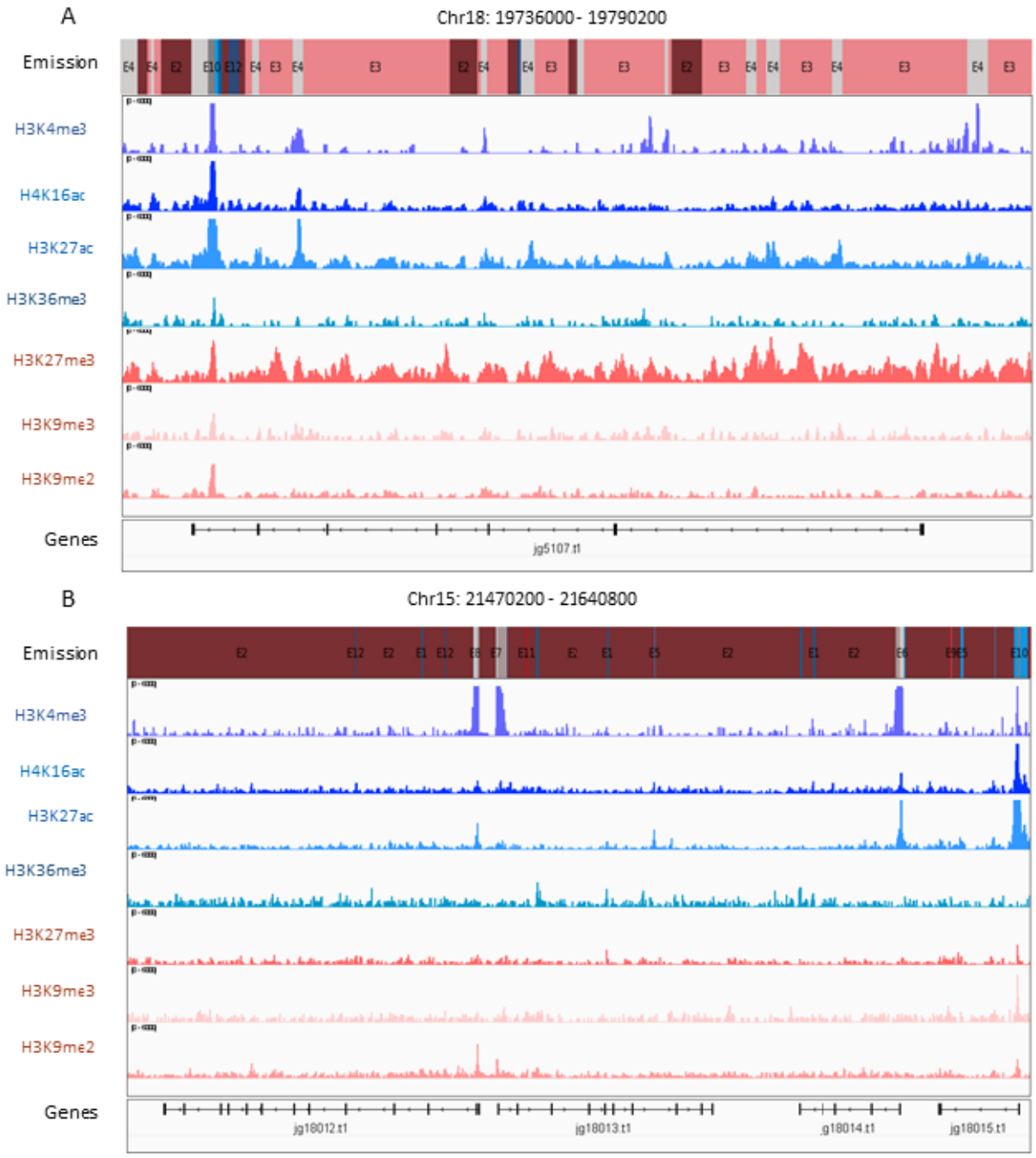
Song, Jimin, and Kevin C. Chen. 2015. "Spectacle: Fast Chromatin State Annotation Using Spectral Learning." *Genome Biology* 16 (1): 33. https://doi.org/10.1186/s13059-015-0598-0.
Straub, Tobias, Charlotte Grimaud, Gregor D. Gilfillan, Angelika Mitterweger, and Peter B. Becker. 2008. "The Chromosomal High-Affinity Binding Sites for the Drosophila Dosage Compensation Complex." *PLOS Genetics* 4 (12): e1000302. https://doi.org/10.1371/journal.pgen.1000302.
Tachibana, Makoto. 2016. "Epigenetics of Sex Determination in Mammals." *Reproductive Medicine and Biology* 15 (2): 59–67. https://doi.org/10.1007/s12522-015-0223-7.
Tenorio, Mariela, Samantha Cruz-Ruiz, Sergio Encarnación-Guevara, Magdalena Hernández, Jose Antonio Corona-Gomez, Fania Sheccid-Santiago, Joanna Serwatowska, et al. 2024. "MAYEX Is an Old Long Noncoding RNA Recruited for X Chromosome Dosage Compensation in a Reptile." *Science* 385 (6715): 1347–54. https://doi.org/10.1126/science.adp1932.
Veitia, Reiner A. 2005. "Gene Dosage Balance: Deletions, Duplications and Dominance." *Trends in Genetics : TIG* 21 (1): 33–35. https://doi.org/10.1016/j.tig.2004.11.002.
Vicoso, Beatriz, J. J. Emerson, Yulia Zektser, Shivani Mahajan, and Doris Bachtrog. 2013. "Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation." *PLoS Biology* 11 (8): e1001643. https://doi.org/10.1371/journal.pbio.1001643.
Walters, James R., Thomas J. Hardcastle, and Chris D. Jiggins. 2015. "Sex Chromosome Dosage Compensation in Heliconius Butterflies: Global yet Still Incomplete?" *Genome Biology and Evolution* 7 (9): 2545–59. https://doi.org/10.1093/gbe/evv156.
Wang, Dapeng, Yubin Zhang, Zhang Zhang, Jiang Zhu, and Jun Yu. 2010. "KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies." *Genomics, Proteomics & Bioinformatics* 8 (1): 77–80. https://doi.org/10.1016/S1672-0229(10)60008-3.
Willforss, Jakob, Aakash Chawade, and Fredrik Levander. 2019. "NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis." *Journal of Proteome Research* 18 (2): 732–40. https://doi.org/10.1021/acs.jproteome.8b00523.
Zhou, Qi, Christopher E. Ellison, Vera B. Kaiser, Artyom A. Alekseyenko, Andrey A. Gorchakov, and Doris Bachtrog. 2013. "The Epigenome of Evolving Drosophila Neo-Sex Chromosomes: Dosage Compensation and Heterochromatin Formation." *PLoS Biology* 11 (11): e1001711. https://doi.org/10.1371/journal.pbio.1001711.
Zimmer F, Fox AM,Pan Q, Ruehle F, Andersen P, Huylmans A-K, Schwander T, Basilicata MF, Keller Valsecchi CI. Convergent Evolution of H4K16ac-mediated Dosage Compensation Shapes Sex-dependent Lifespan in a ZW Species. *bioRxiv*. 2025. 10.1101/2025.04.10.648168



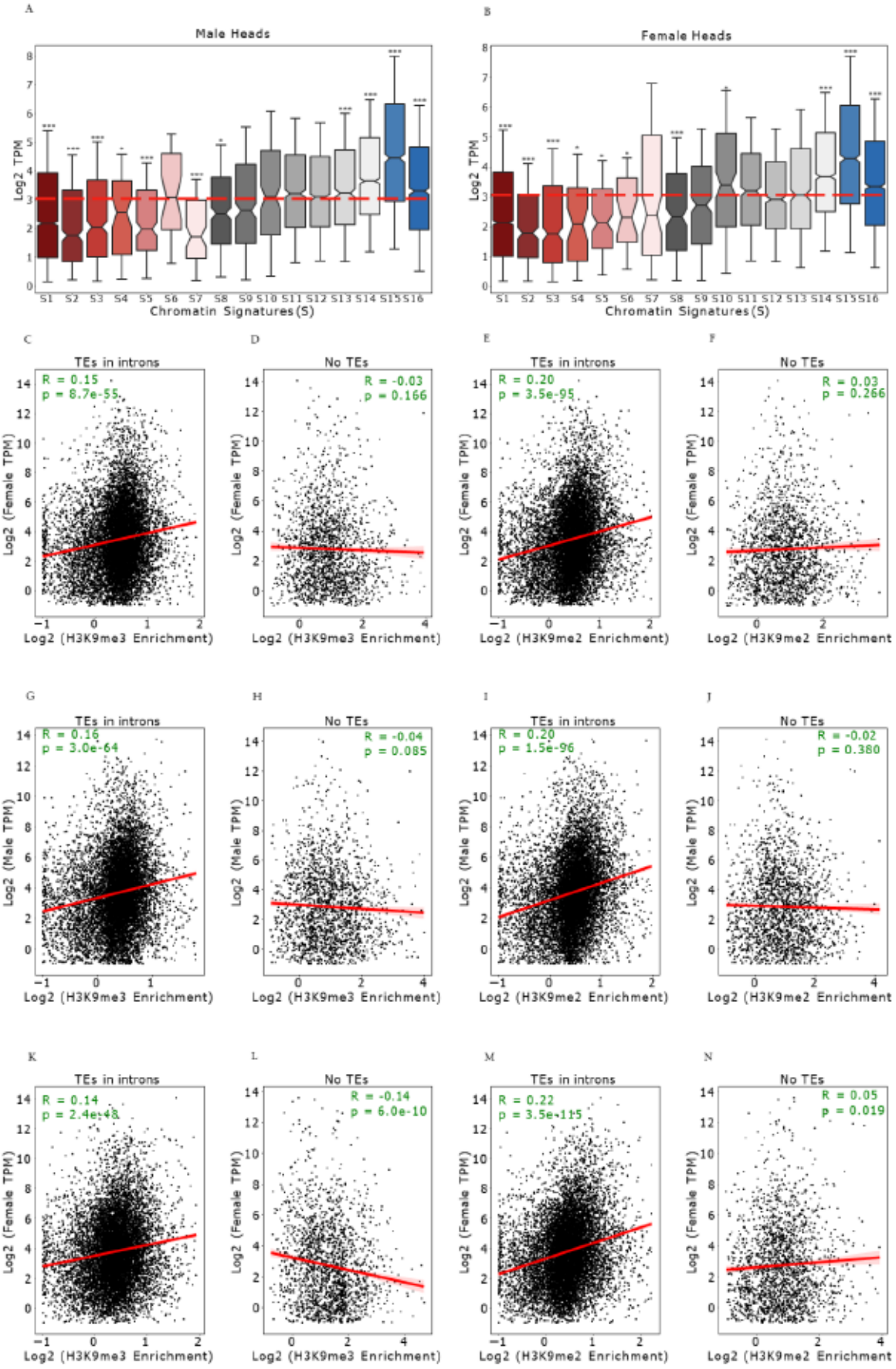
Supplementary Figure 1: Pearson Correlation of selected model (E21) to sets of models (E20:E2) to identify optimal number of chromatin states.



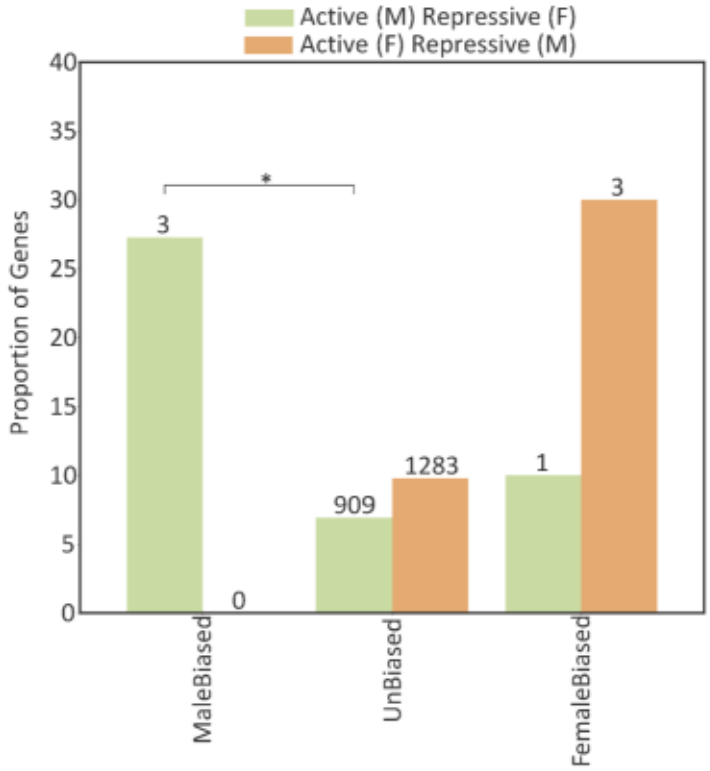
Supplementary Figure 2: Emission model showing likelihood of finding histone mark in each chromatin state number. The darker the color means higher chance of presence of that particular histone mark.



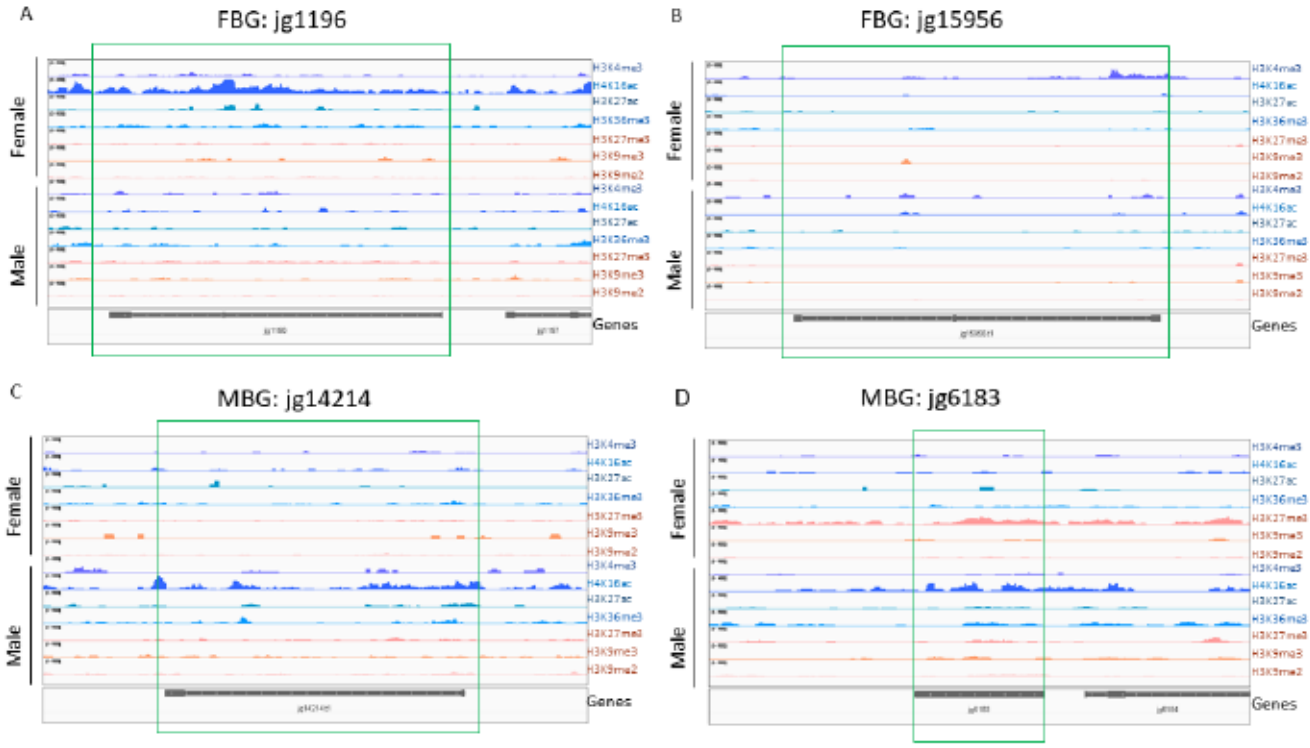
Supplementary Figure 3: Representative emission state profiles for genomic regions of chromosome 18 (A) and chromosome 15 (B) in male head tissue.


A
Male Heads
Log2 TPM
Chromatin Signatures (S)
B
Female Heads
Log2 TPM
Chromatin Signatures (S)
C
TEs in introns
R = 0.15
p = 8.7e-55
Log2 (Female TPM)
Log2 (H3K9me3 Enrichment)
D
No TEs
R = -0.03
p = 0.166
Log2 (Female TPM)
Log2 (H3K9me3 Enrichment)
E
TEs in introns
R = 0.20
p = 3.5e-95
Log2 (Female TPM)
Log2 (H3K9me2 Enrichment)
F
No TEs
R = 0.03
p = 0.266
Log2 (Female TPM)
Log2 (H3K9me2 Enrichment)
G
TEs in introns
R = 0.16
p = 3.0e-64
Log2 (Male TPM)
Log2 (H3K9me3 Enrichment)
H
No TEs
R = -0.04
p = 0.085
Log2 (Male TPM)
Log2 (H3K9me3 Enrichment)
I
TEs in introns
R = 0.20
p = 1.5e-96
Log2 (Male TPM)
Log2 (H3K9me2 Enrichment)
J
No TEs
R = -0.02
p = 0.380
Log2 (Male TPM)
Log2 (H3K9me2 Enrichment)
K
TEs in introns
R = 0.14
p = 2.4e-48
Log2 (Female TPM)
Log2 (H3K9me3 Enrichment)
L
No TEs
R = -0.14
p = 6.0e-10
Log2 (Female TPM)
Log2 (H3K9me3 Enrichment)
M
TEs in introns
R = 0.22
p = 3.5e-115
Log2 (Female TPM)
Log2 (H3K9me2 Enrichment)
N
No TEs
R = 0.05
p = 0.019
Log2 (Female TPM)
Log2 (H3K9me2 Enrichment)

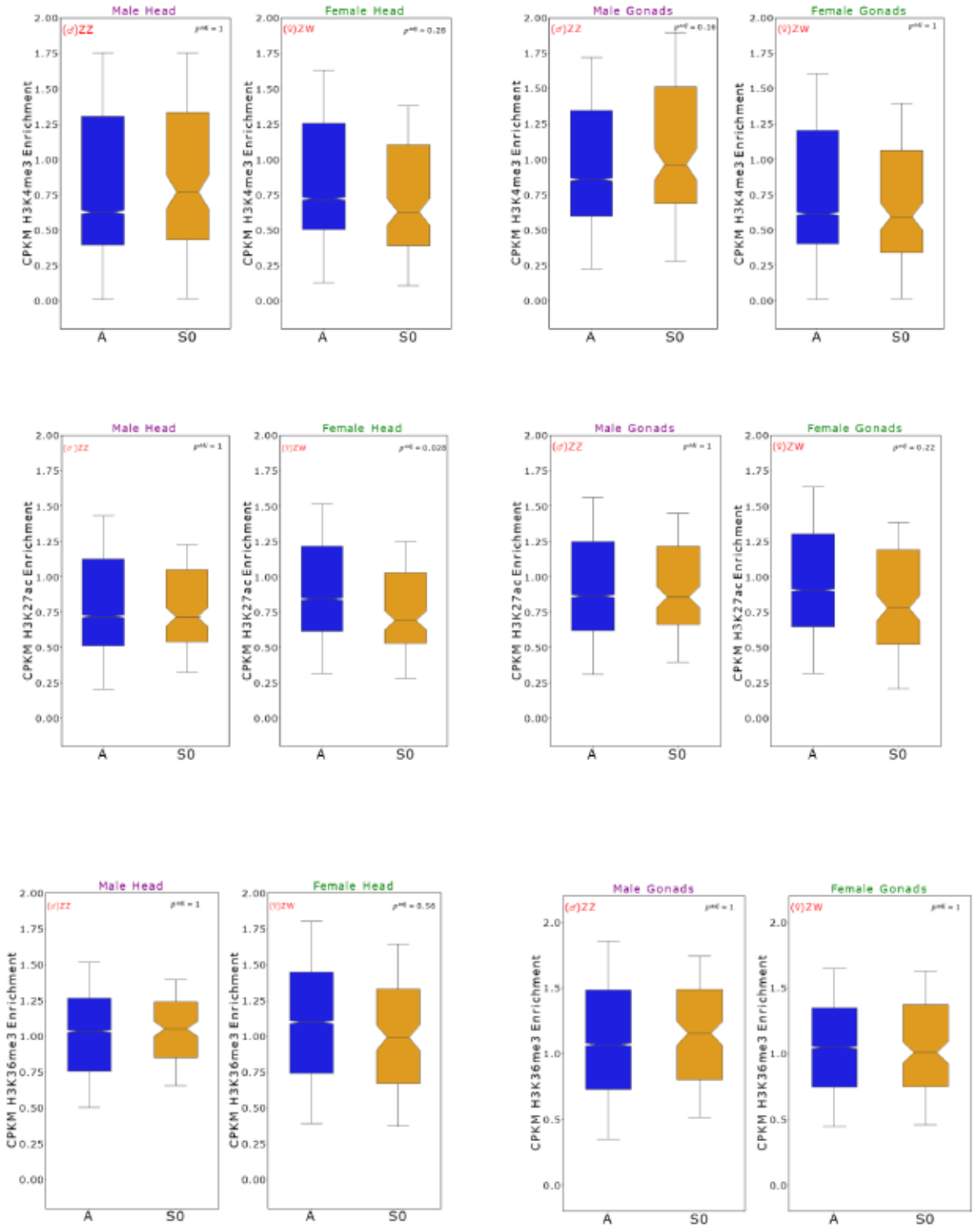
Supplementary Figure 4: Expression of chromatin signatures (W and Z-specific genes were excluded) in male (A) and female heads (B). Panels C to D show the correlation of female heads' expression with H3K9me3 for genes with (C) or without (D) TEs in their introns. Panels G and H show the same for male gonads expression. Panels K and L show the same for female gonads expression. Panels E and F show the correlation of female heads' expression with H3K9me2 for genes with (E) or without (F) TEs in their introns. Panels I and J show the same for male gonads expression. Panels M and N show the same for female gonads expression.



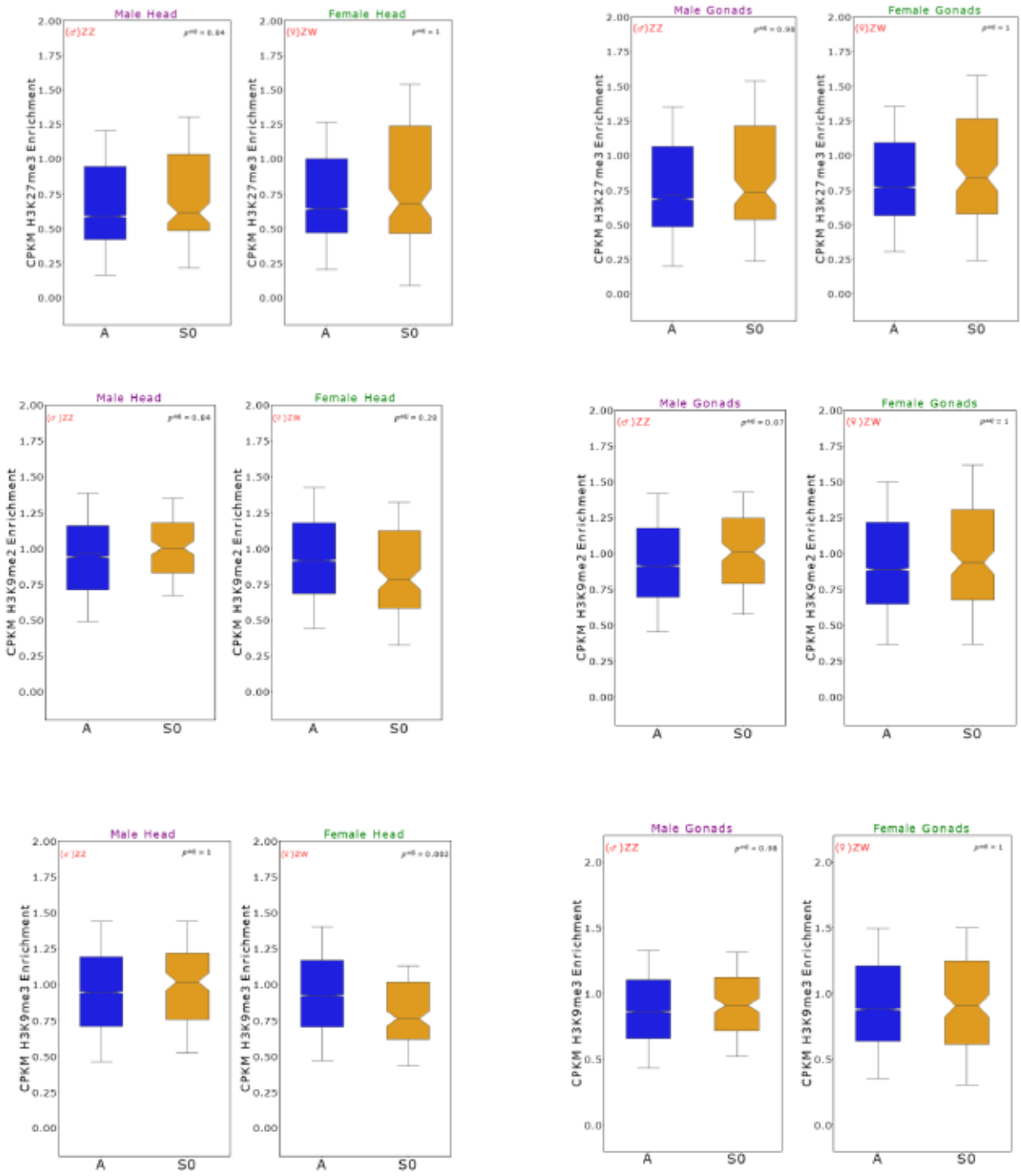
Supplementary Figure 5: Proportion of genes with contrasting chromatin states in males and females and which have sex-biased expression in heads of *A. franciscana* (FDR <0.05, Foldchange >2 & TPM >0.1 for genes in *A. franciscana*)



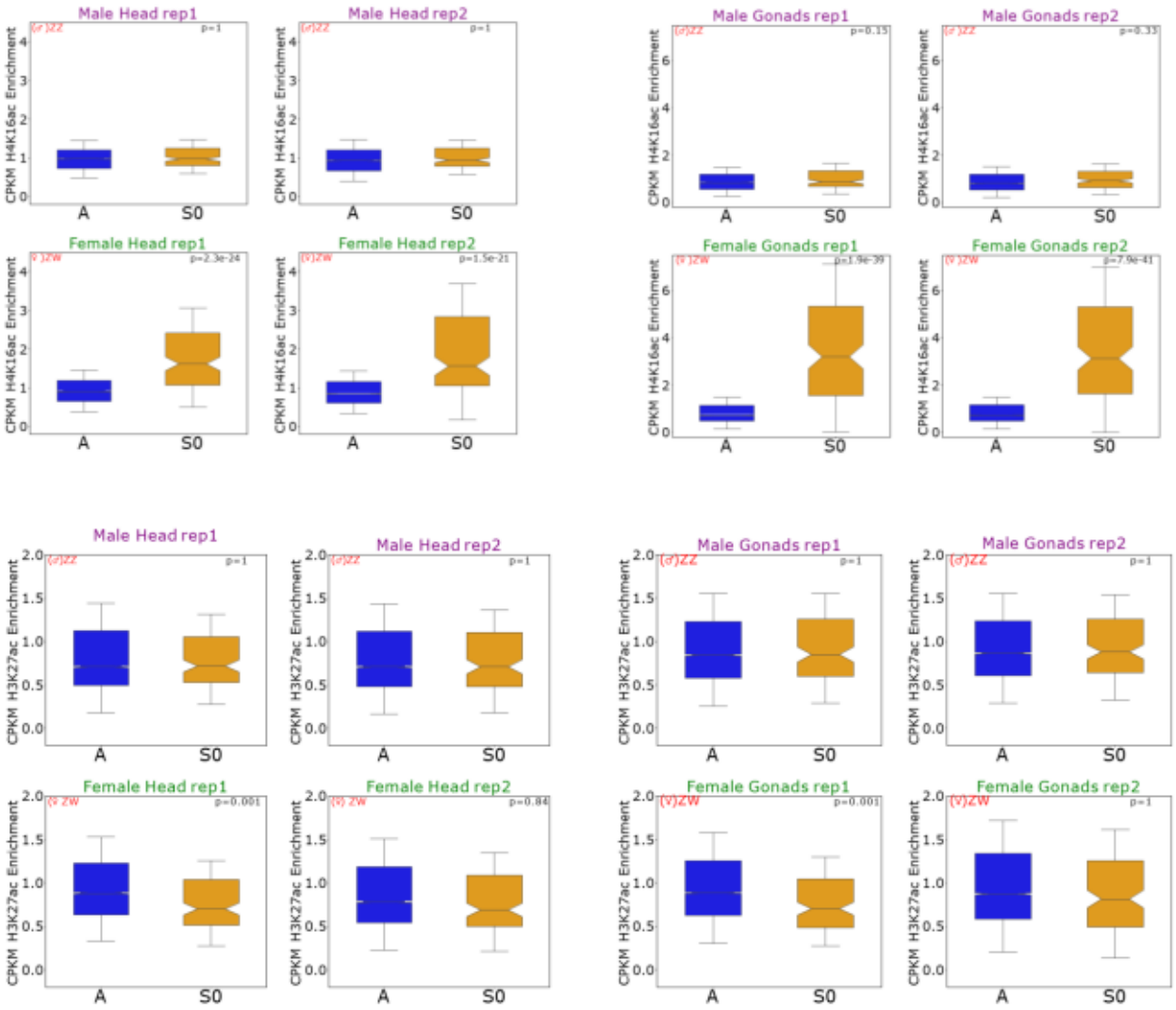
Supplementary Figure 6: **Representative chromatin profile regions for female-biased genes** (A and B) on chromosome 20 (jg1196; is enriched with H4K16ac (E12, chromatin signature S16) in females and null states in males E2 (S1) and chromosome 17 (jg15956; is enriched with H3K4me3 (E7, chromatin signature S16) in females and null states E2 (S1) in males). **Representative chromatin profile regions for male-biased genes** (C and D) on chromosome 7 (jg14214; is enriched with H4K16ac (E12 chromatin signature S16) in males and null states in females E2 (chromatin signature S1) and chromosome 16 (jg6183; is enriched with H4K16ac (E12, chromatin signature S16 in males) while in female is enriched with H3K27me3 (E3, chromatin signature S2) .

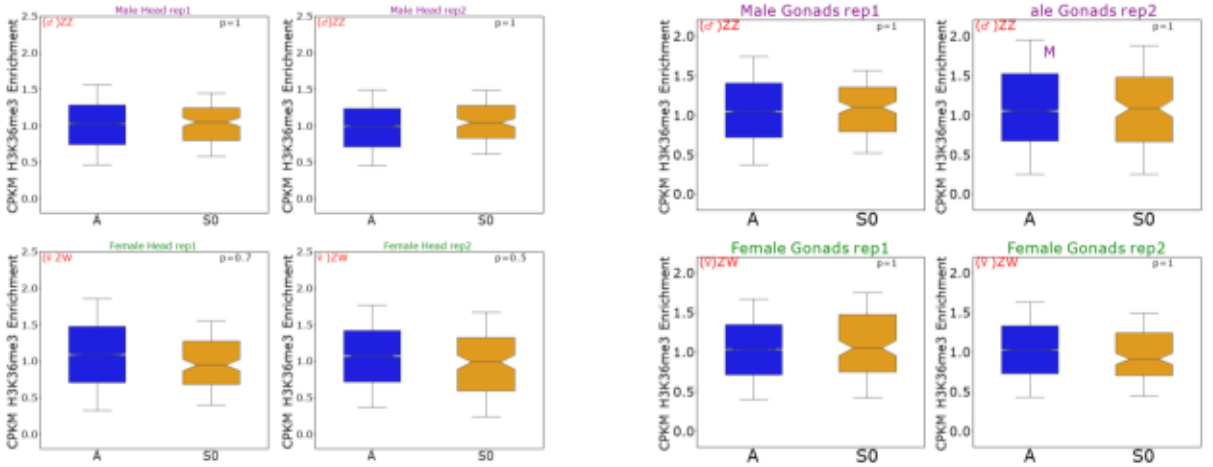


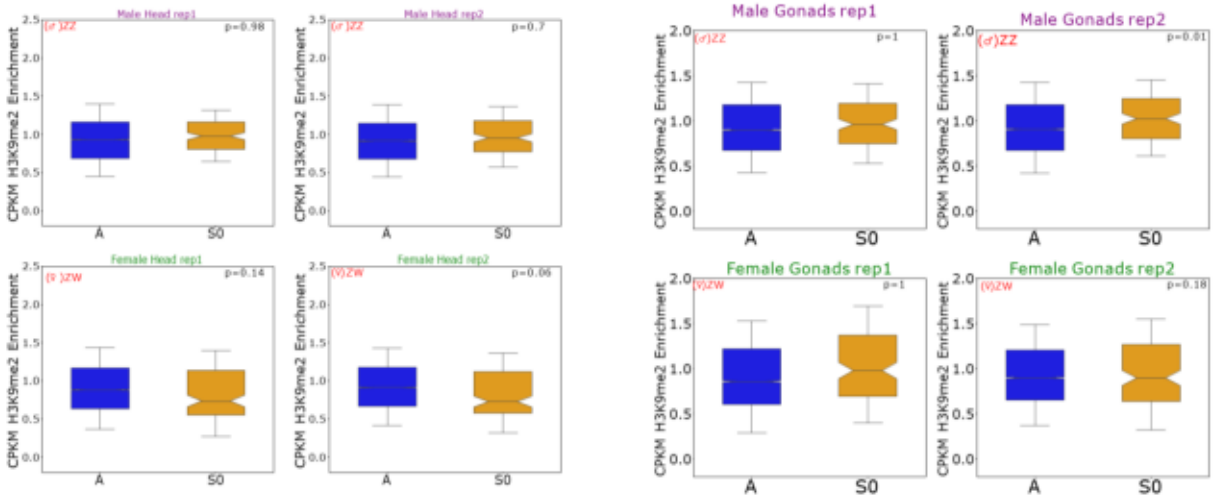
Supplementary Figure 7: Normalized CPKM enrichment of active associated chromatin marks based on autosomal and Z-specific genes in somatic and gonadal tissues of *A. franciscana*.

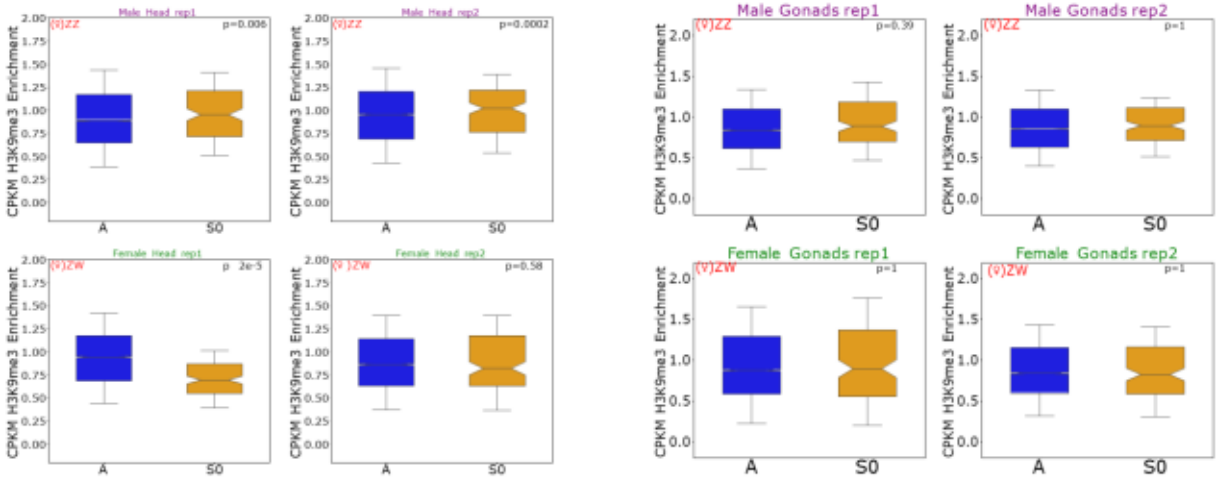


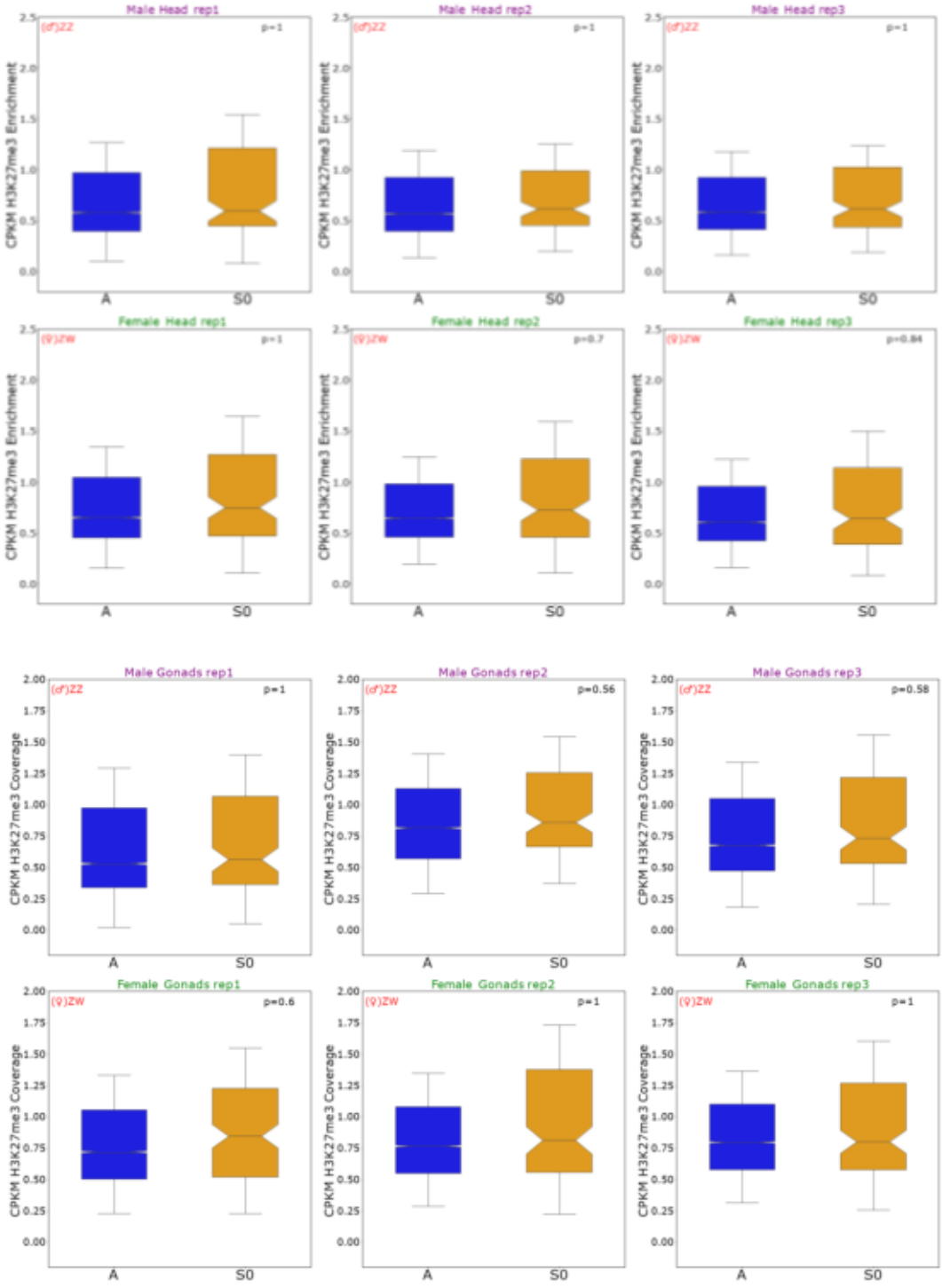
Supplementary Figure 8: Normalized CPKM enrichment of repressive associated chromatin marks based on autosomal and Z-specific genes in somatic and gonadal tissues of *A. franciscana*.

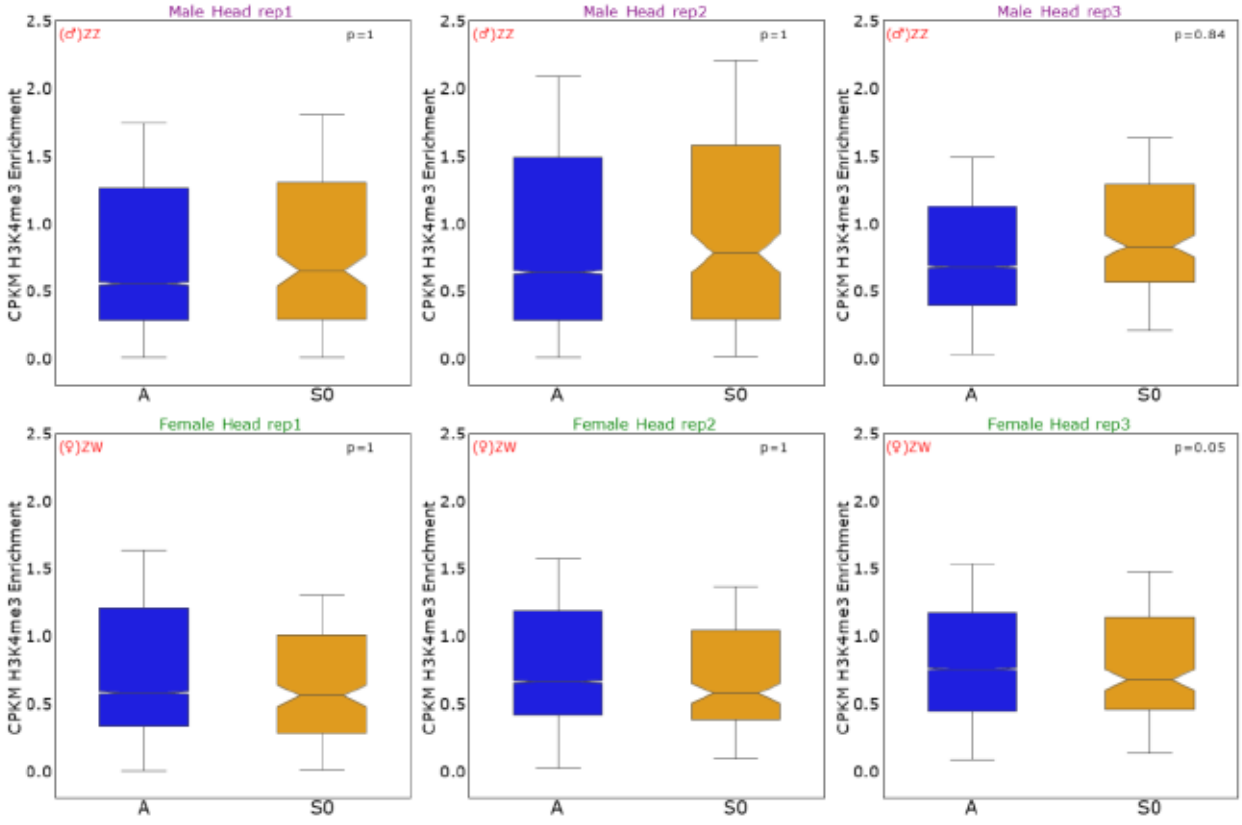

Male Head rep1
Male Head rep2
Male Gonads rep1
Male Gonads rep2
Female Head rep1
Female Head rep2
Female Gonads rep1
Female Gonads rep2
CPKM H4K16ac Enrichment
CPKM H3K27ac Enrichment
A
S0

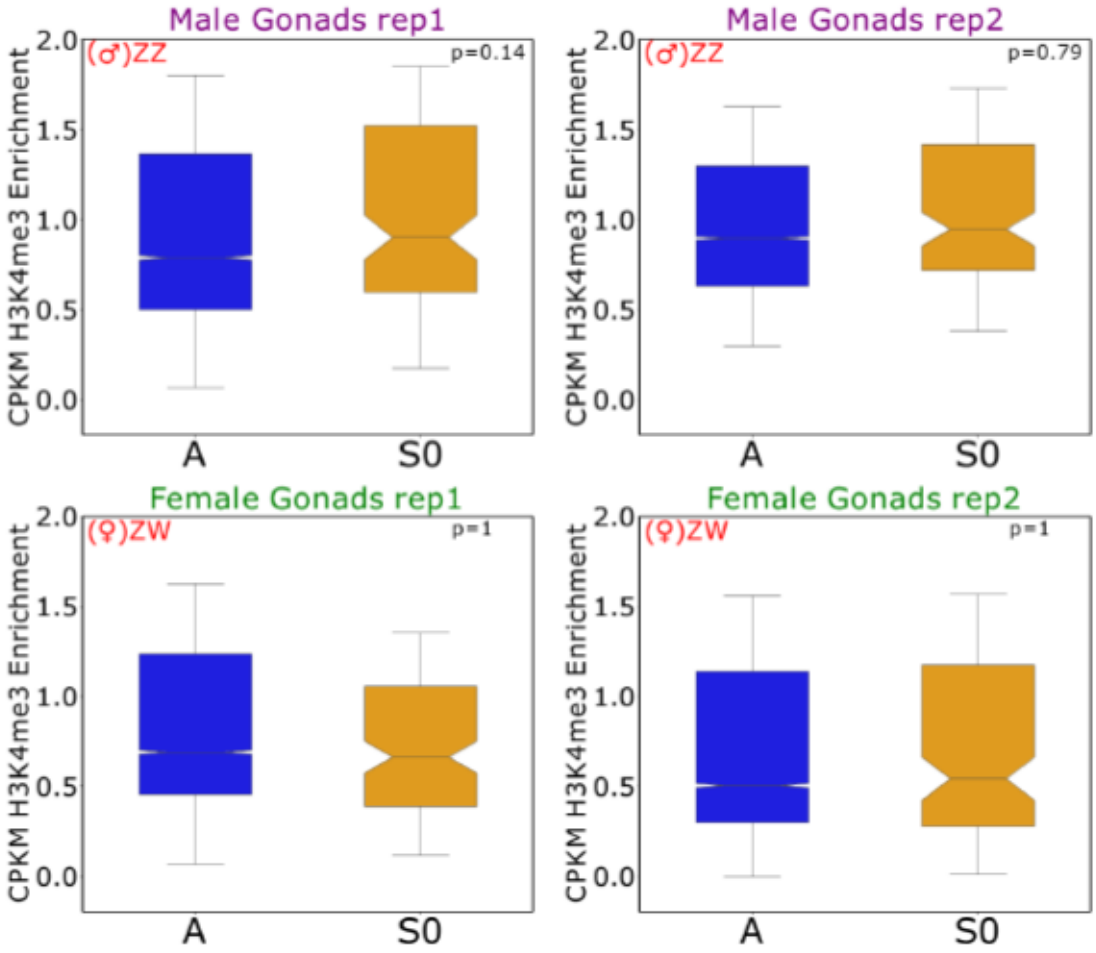

Male Head rep1
Male Head rep2
Male Gonads rep1
ale Gonads rep2
Female Head rep1
Female Head rep2
Female Gonads rep1
Female Gonads rep2
CPKM H3K36me3 Enrichment
A
S0


Male Head rep1
Male Head rep2
Male Gonads rep1
Male Gonads rep2
Female Head rep1
Female Head rep2
Female Gonads rep1
Female Gonads rep2
CPKM H3K9me2 Enrichment
A
S0

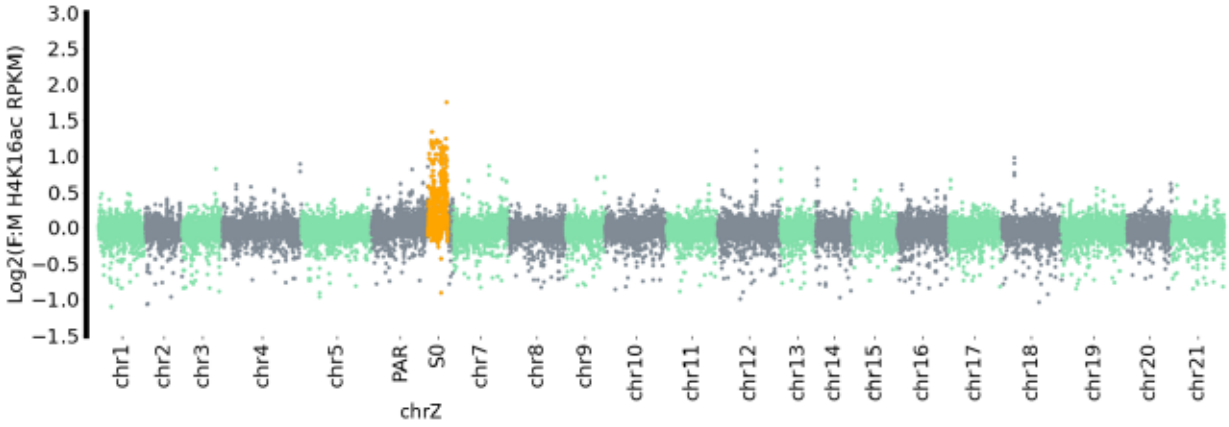

Male Head rep1
Male Head rep2
Male Gonads rep1
Male Gonads rep2
Female Head rep1
Female Head rep2
Female Gonads rep1
Female Gonads rep2
CPKM H3K9me3 Enrichment
A
S0


Male Head rep1
Male Head rep2
Male Head rep3
Female Head rep1
Female Head rep2
Female Head rep3
Male Gonads rep1
Male Gonads rep2
Male Gonads rep3
Female Gonads rep1
Female Gonads rep2
Female Gonads rep3
CPKM H3K27me3 Enrichment
CPKM H3K27me3 Coverage
A
S0

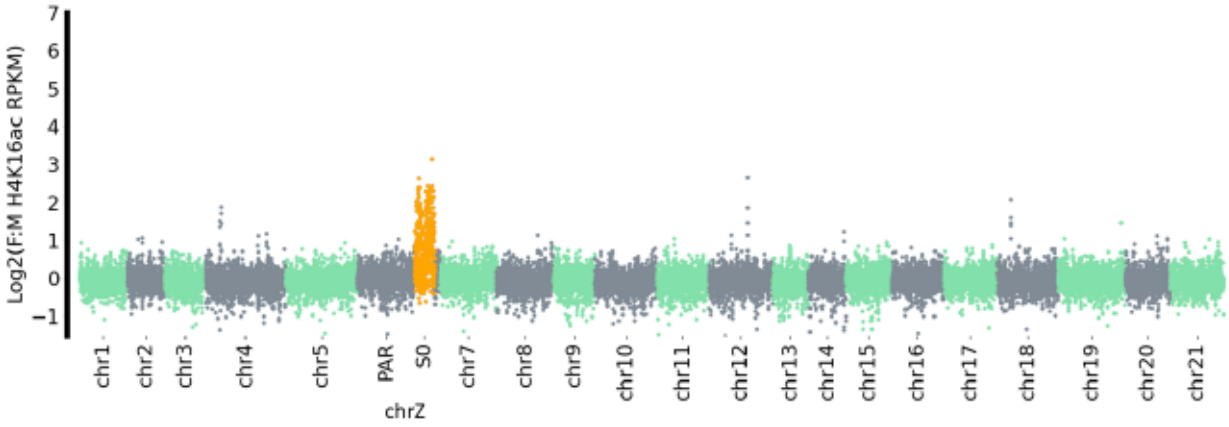

Male Head rep1
(♂)ZZ
p=1
Male Head rep2
(♂)ZZ
p=1
Male Head rep3
(♂)ZZ
p=0.84
Female Head rep1
(♀)ZW
p=1
Female Head rep2
(♀)ZW
p=1
Female Head rep3
(♀)ZW
p=0.05
CPKM H3K4me3 Enrichment
0.0
0.5
1.0
1.5
2.0
2.5
A
S0



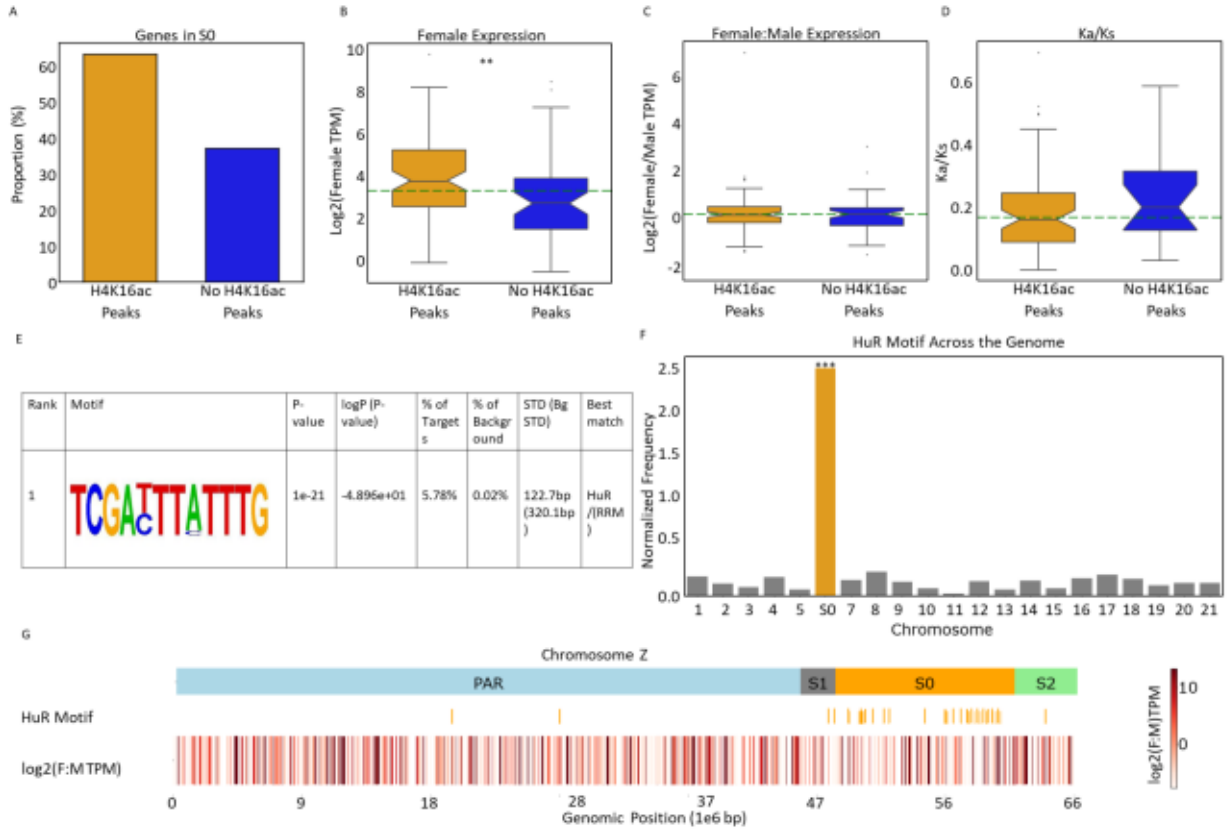
Supplementary Figure 9: Normalized CPKM enrichment of both active and repressive associated chromatin marks based on autosomal and Z-specific genes in each replicate in head and gonadal tissues of *A. franciscana*



Supplementary Figure 10: Female to male coverage distribution of H4K16ac in 30kb windows across 21 chromosomes in Heads in *A. franciscana*.

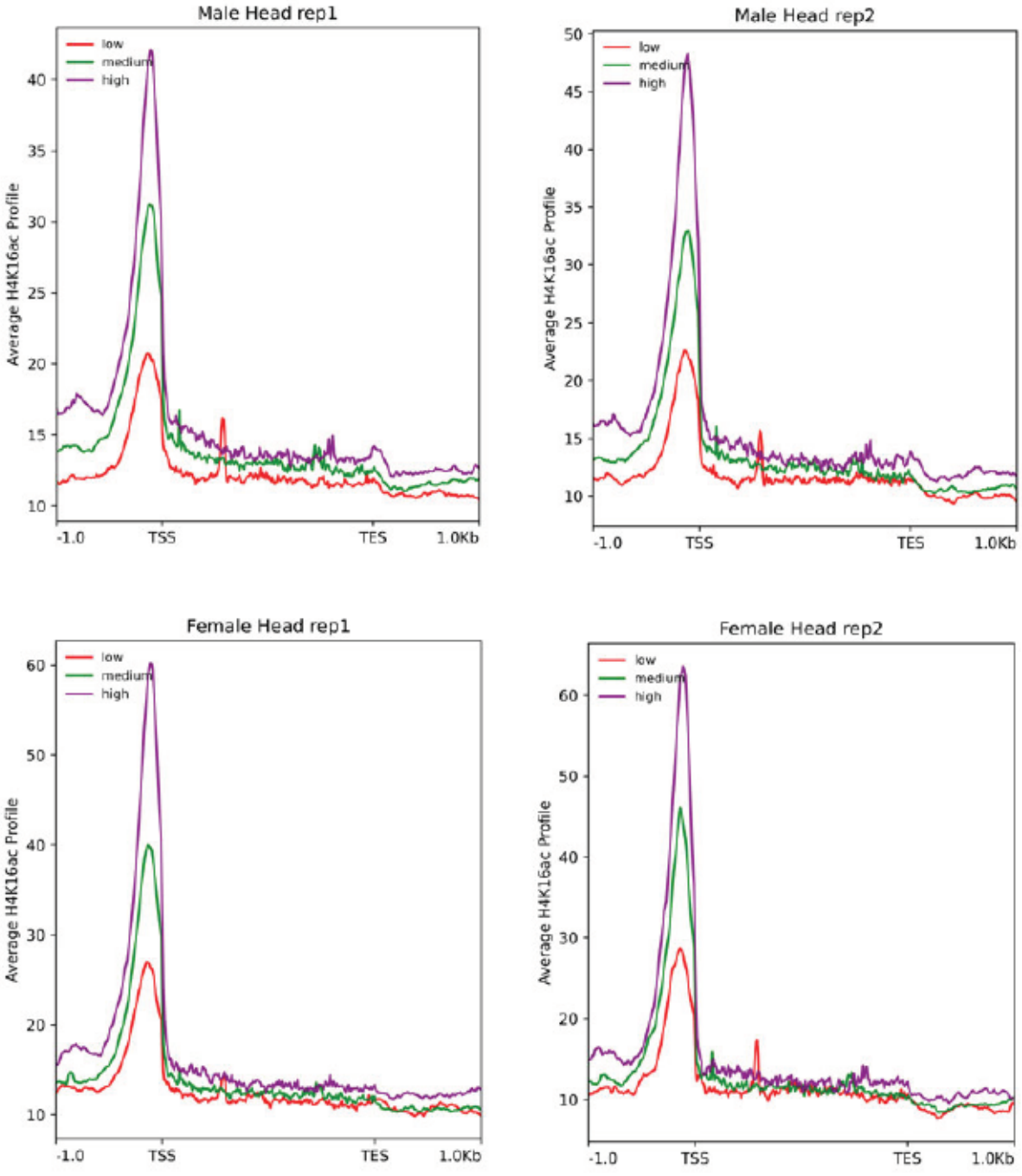


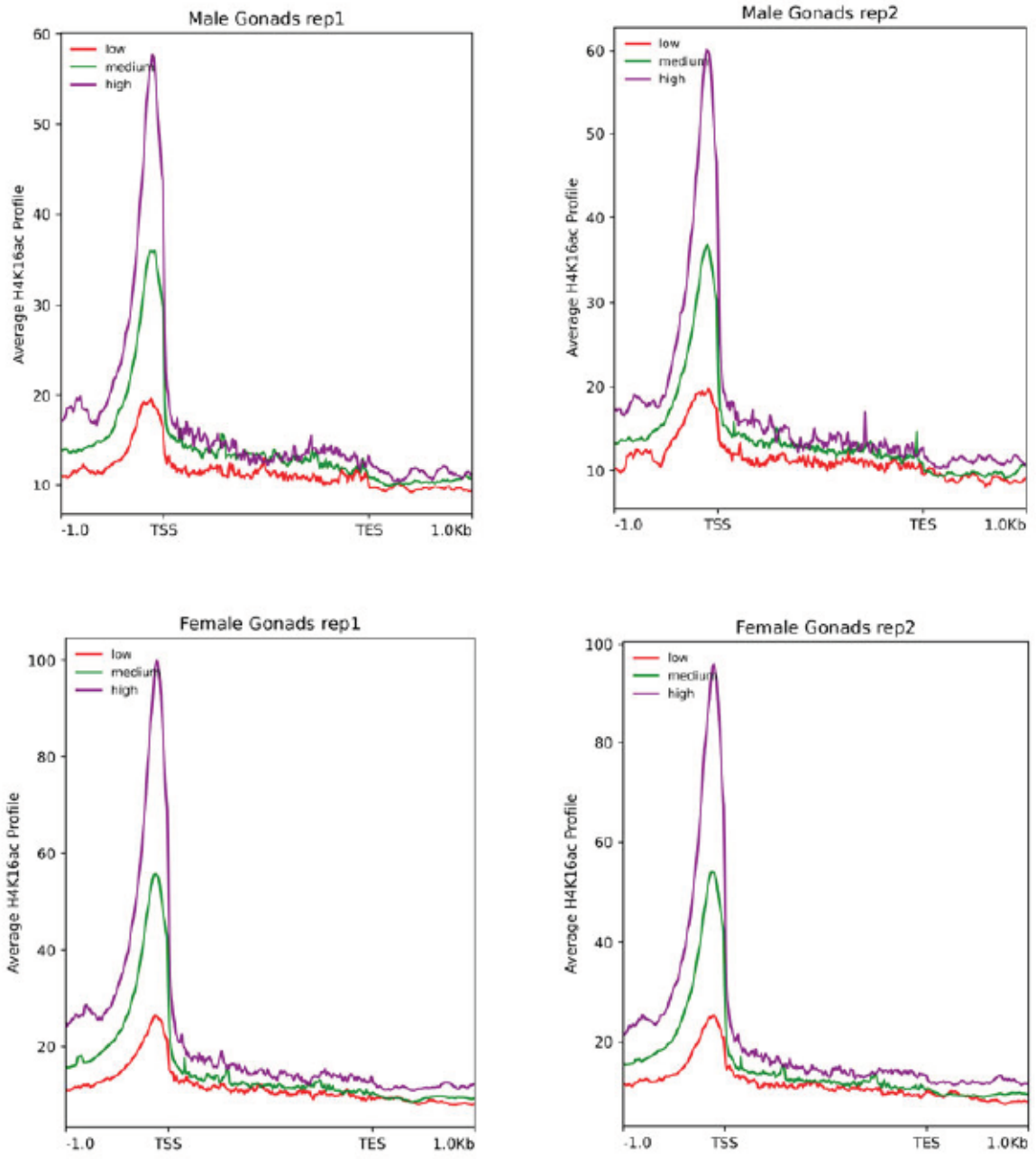
Supplementary Figure 11: Female to male coverage distribution of H4K16ac in 30kb windows across 21 chromosomes in Gonads in *A. franciscana*.



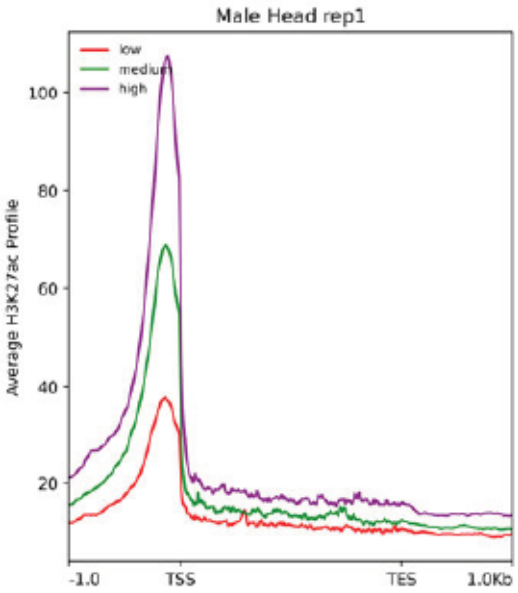
Supplementary Figure 12: **H4K16ac peaks in the S0 region and their association with functional genomic features.** (A) Proportion of genes in the Z-specific region (S0) that either intersect or do not intersect with H4K16ac peaks in female gonads. (B) Expression of genes in the Z-specific region (S0) with and without H4K16ac peaks in female gonads. The significance of Wilcoxon rank sum test ** is *P*-value < 0.005. (C) Log2 of female:male expression ratio of S0 genes with and without H4K16ac peaks in gonads. (D) The distribution of Ka/Ks of S0 genes with and without overlap with H4K16ac peaks. (E) Motif with the most significant enrichment in S0 H4K16ac peaks relative to the rest of the genome. (F) Frequency (in number per million base pairs) of the motif described in (E) on different chromosomes. *** denotes

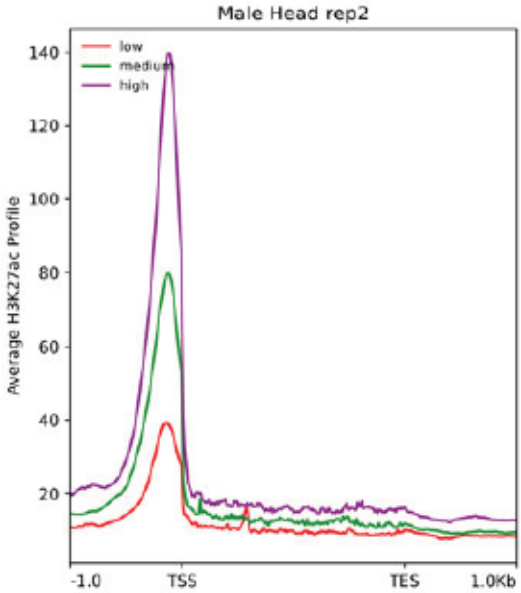
p<0.005, obtained by resampling N chromosomal loci randomly 10000 times, where N is the number of peaks found throughout the genome. (G) Schematic regions of the Z chromosome, HuR motif distribution patterns and their relationship to female-to-male expression ratios in gonadal tissue.

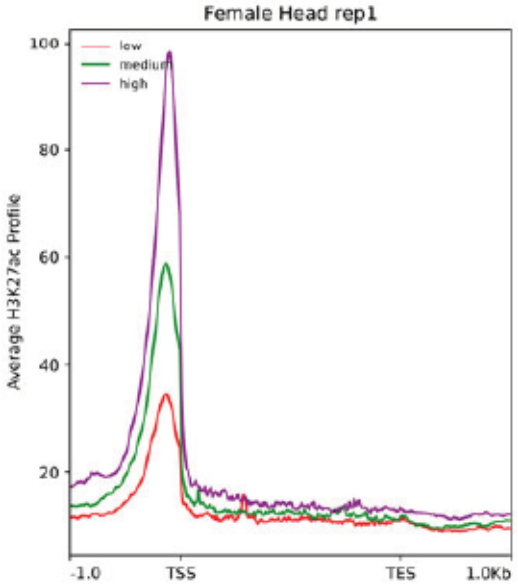


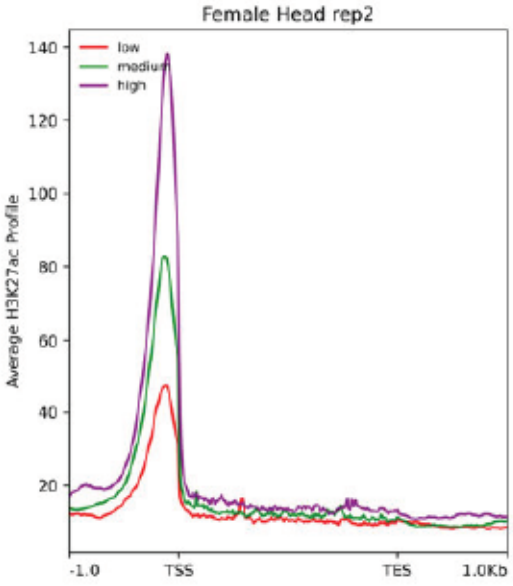


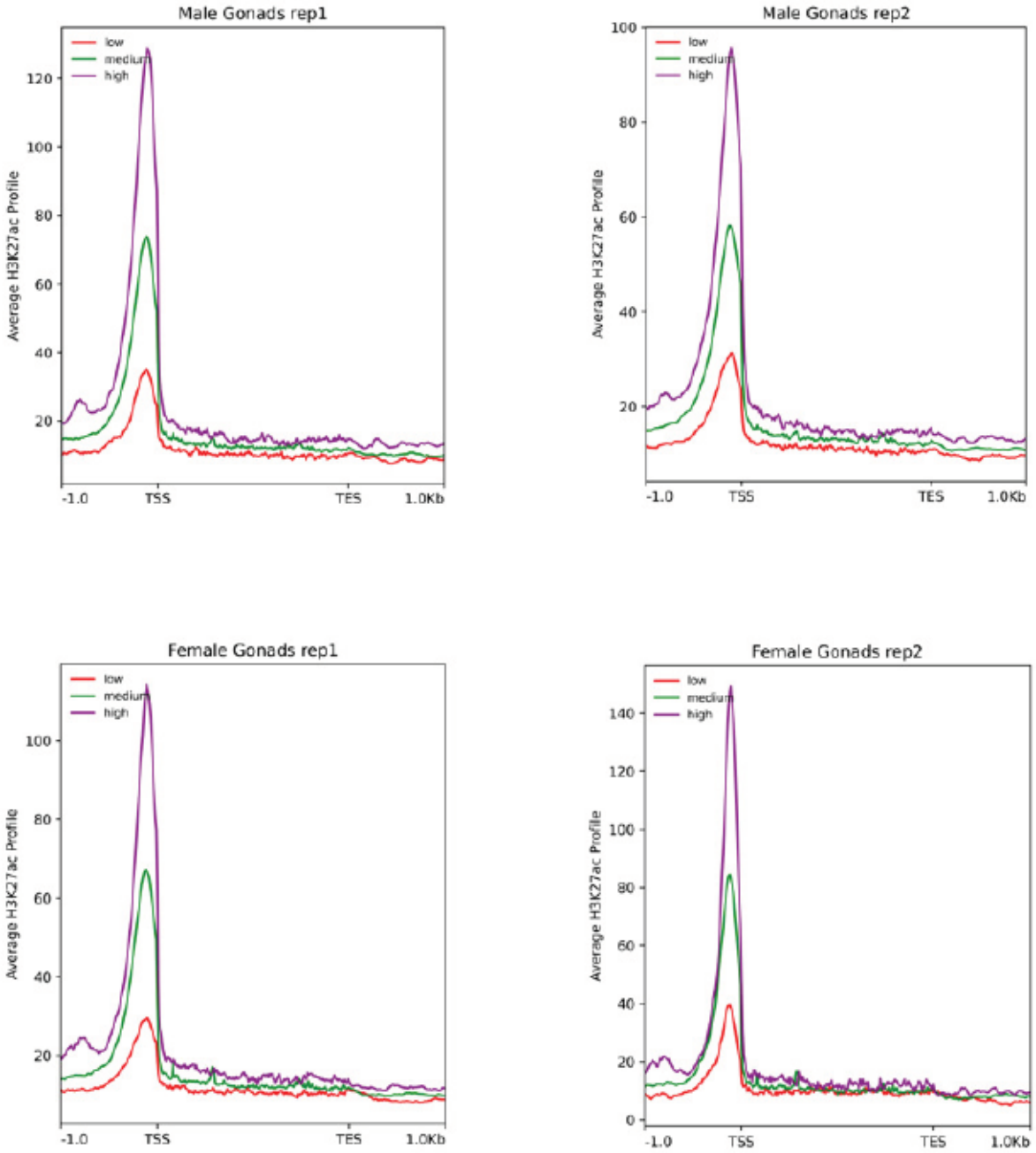
Supplementary Figure 13: Average enrichment of H4K16ac across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads.


Male Head rep1
low
medium
high
Average H3K27ac Profile
100
80
60
40
20
-1.0
TSS
TES
1.0Kb

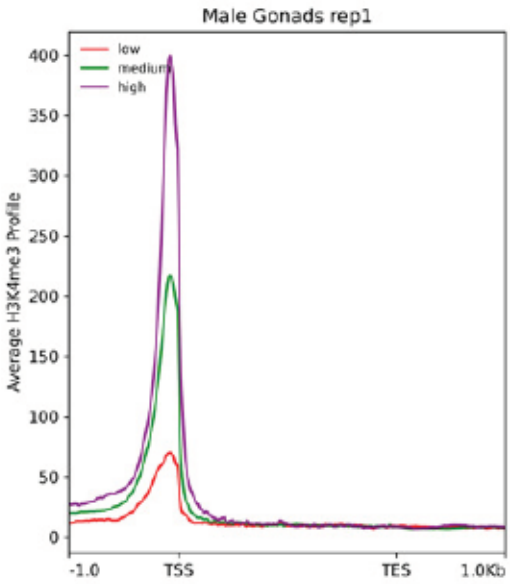

Male Head rep2
low
medium
high
Average H3K27ac Profile
140
120
100
80
60
40
20
-1.0
TSS
TES
1.0Kb

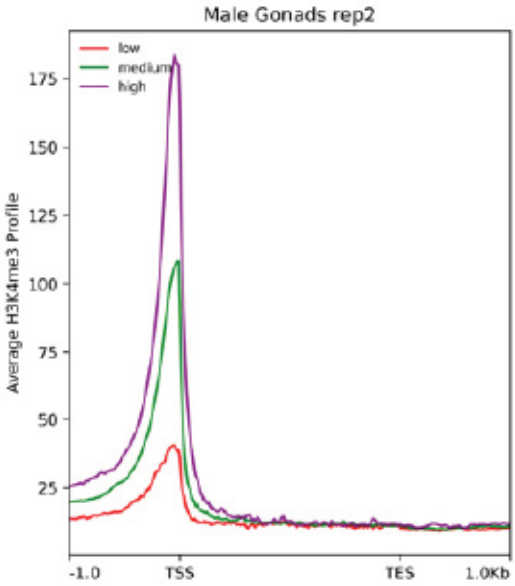

Female Head rep1
low
medium
high
Average H3K27ac Profile
100
80
60
40
20
-1.0
TSS
TES
1.0Kb

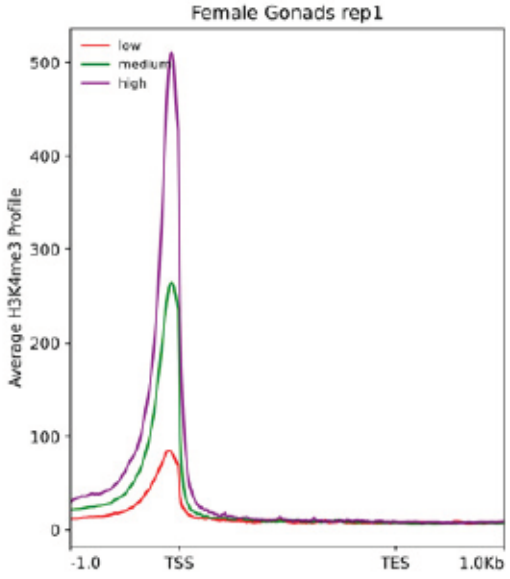

Female Head rep2
low
medium
high
Average H3K27ac Profile
140
120
100
80
60
40
20
-1.0
TSS
TES
1.0Kb

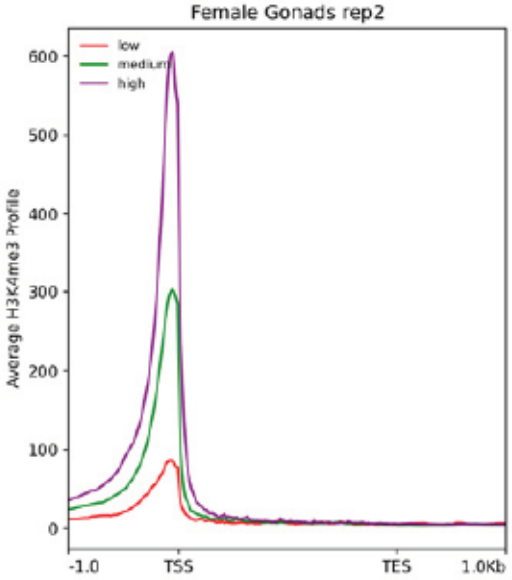


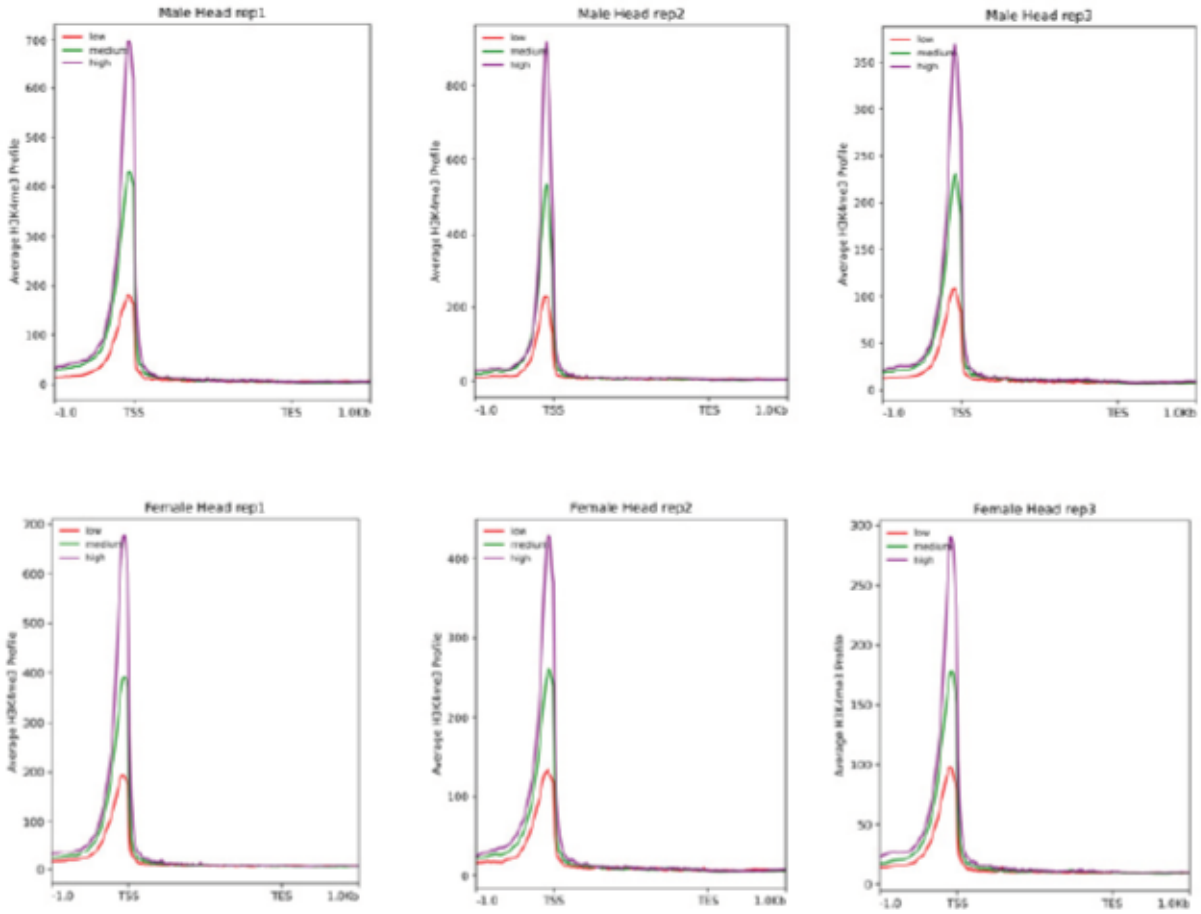
Supplementary Figure 14: Average enrichment of H3K27ac across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads.


Male Gonads rep1
low
medium
high
Average H3K4me3 Profile
400
350
300
250
200
150
100
50
0
-1.0
TSS
TES
1.0Kb

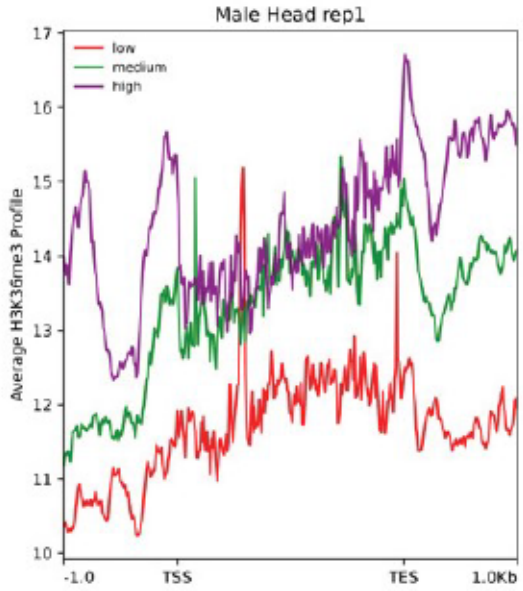

Male Gonads rep2
low
medium
high
Average H3K4me3 Profile
175
150
125
100
75
50
25
-1.0
TSS
TES
1.0Kb

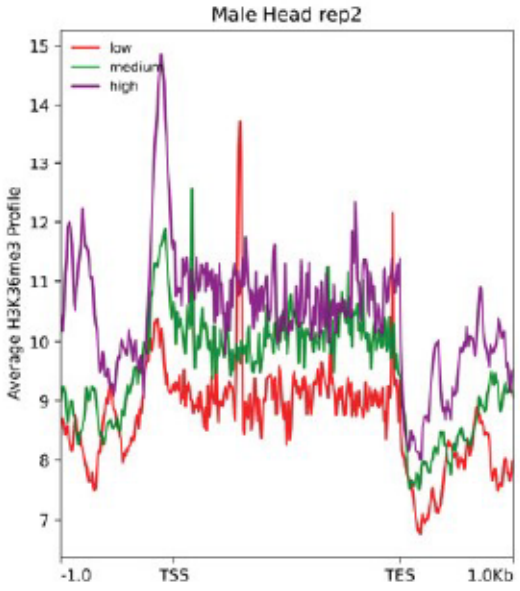

Female Gonads rep1
low
medium
high
Average H3K4me3 Profile
500
400
300
200
100
0
-1.0
TSS
TES
1.0Kb

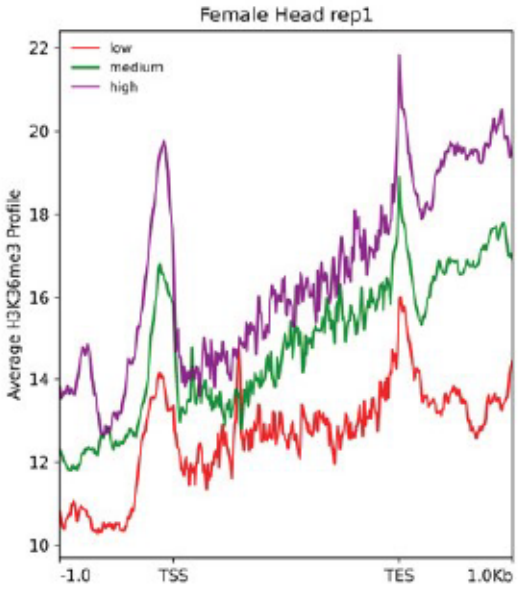

Female Gonads rep2
low
medium
high
Average H3K4me3 Profile
600
500
400
300
200
100
0
-1.0
TSS
TES
1.0Kb

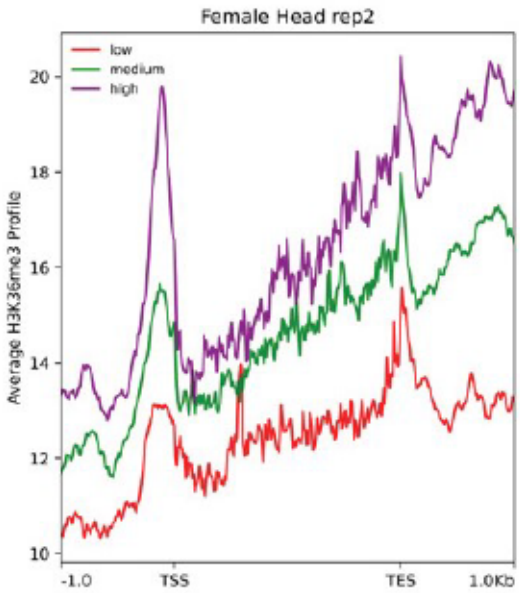


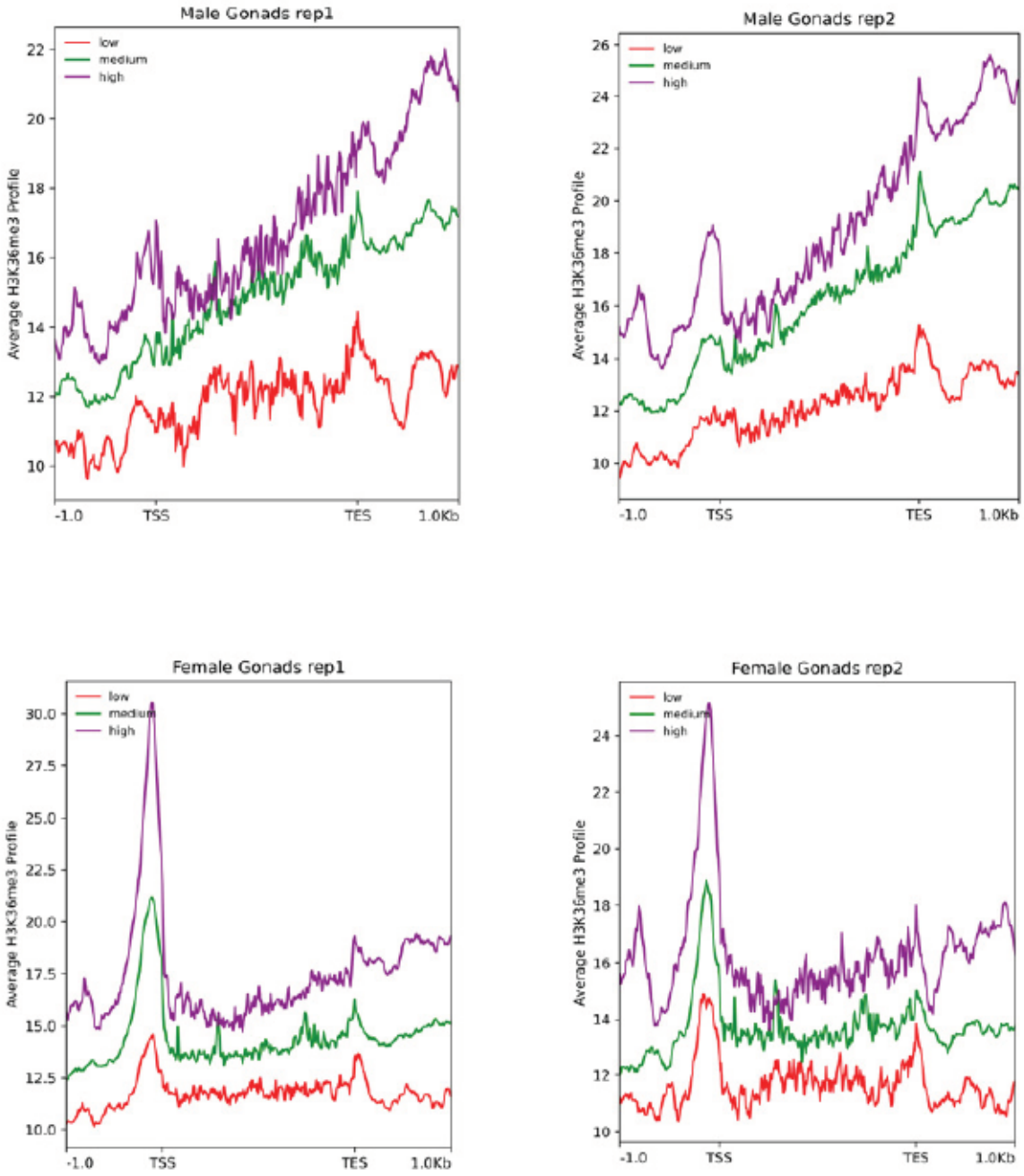
Supplementary Figure 15: Average enrichment of H3K4me3 across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads.


Male Head rep1
low
medium
high
Average H3K36me3 Profile
-1.0
TSS
TES
1.0Kb

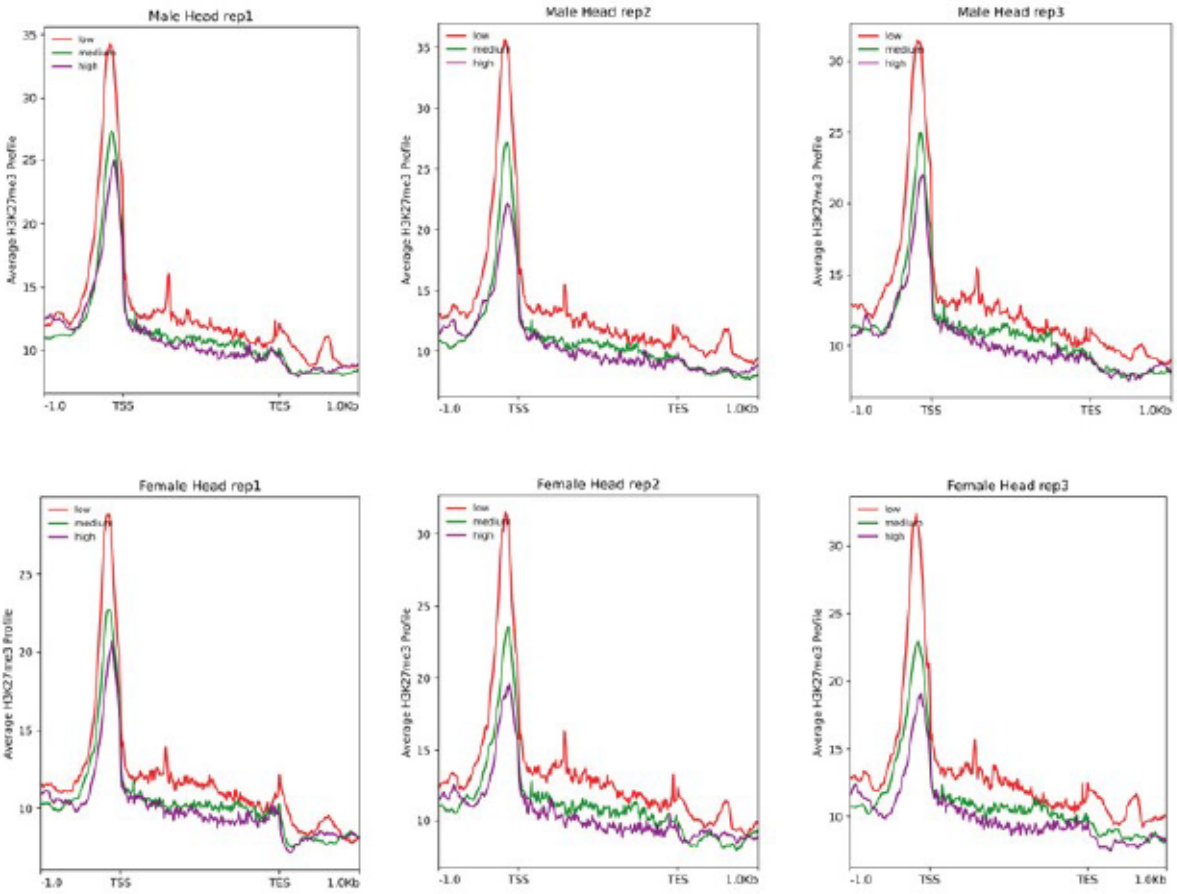

Male Head rep2
low
medium
high
Average H3K36me3 Profile
-1.0
TSS
TES
1.0Kb

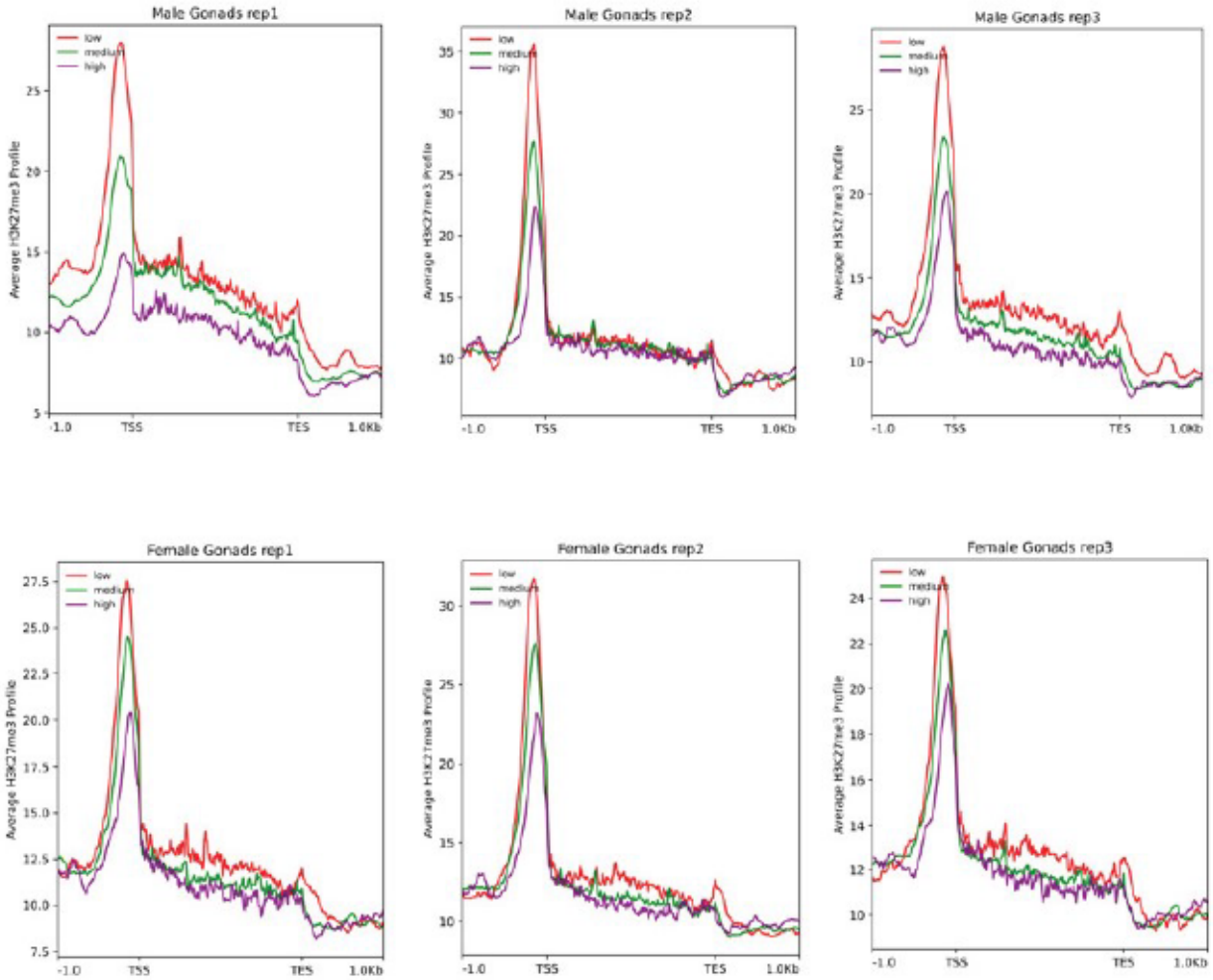

Female Head rep1
low
medium
high
Average H3K36me3 Profile
-1.0
TSS
TES
1.0Kb


Female Head rep2
low
medium
high
Average H3K36me3 Profile
-1.0
TSS
TES
1.0Kb

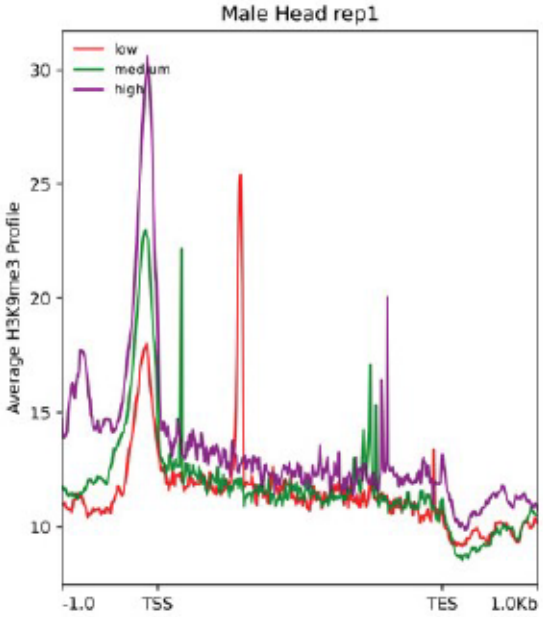


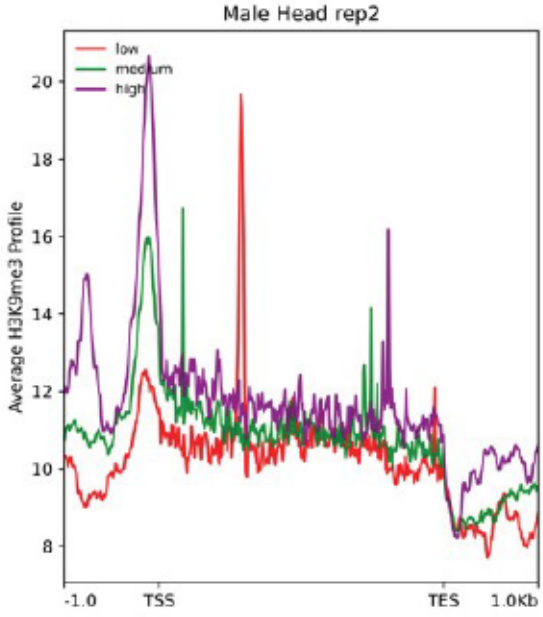
Supplementary Figure 16: Average enrichment of H3K36me3 across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads

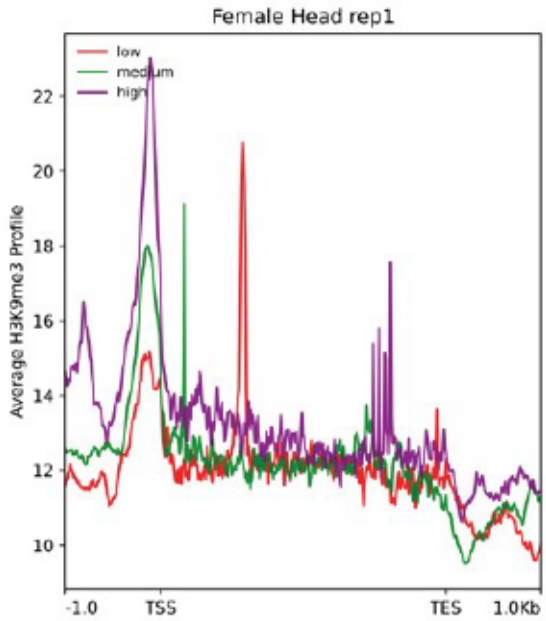

Male Head rep1
Male Head rep2
Male Head rep3
Female Head rep1
Female Head rep2
Female Head rep3
low
medium
high
Average H3K27me3 Profile
-1.0
TSS
TES
1.0Kb

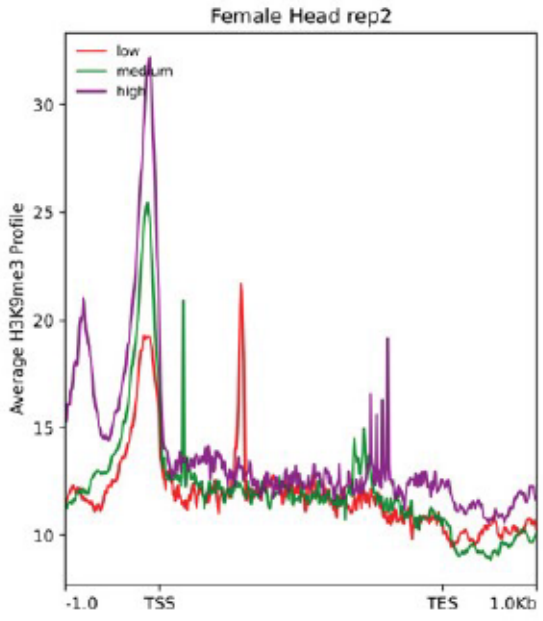


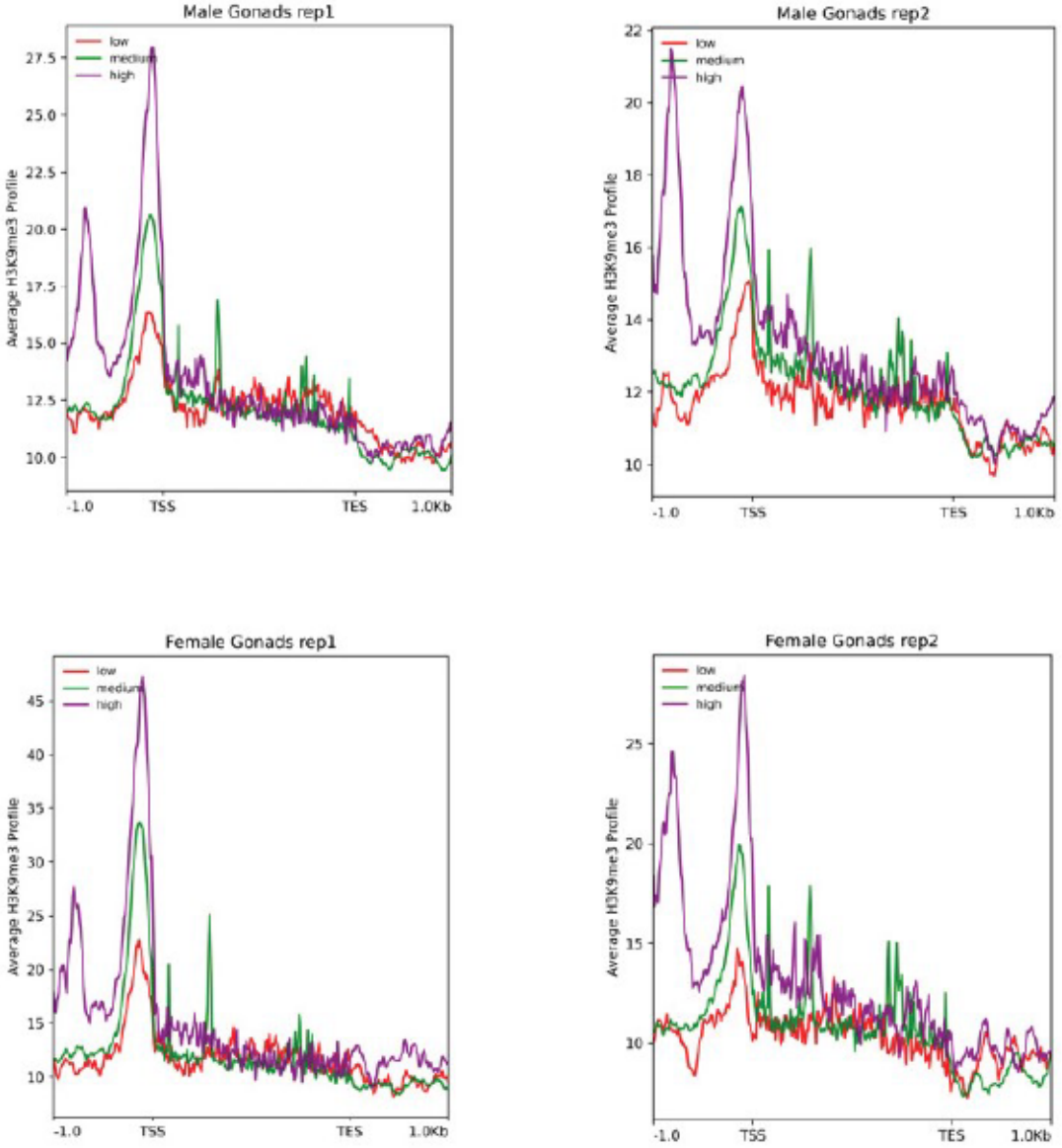
Supplementary Figure 17: Average enrichment of H3K27me3 across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads


Male Head rep1
low
medium
high
Average H3K9me3 Profile
30
25
20
15
10
-1.0
TSS
TES
1.0Kb

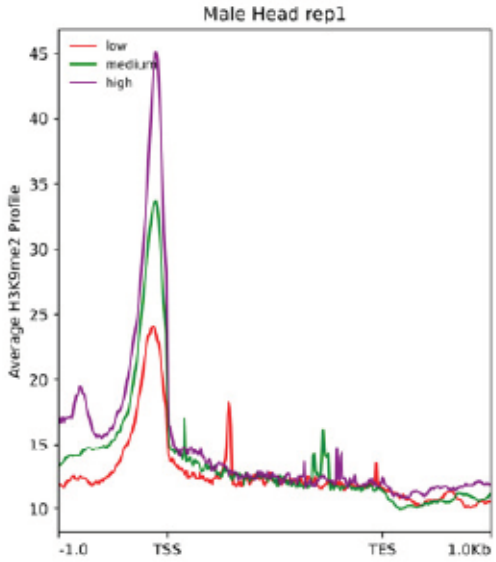

Male Head rep2
low
medium
high
Average H3K9me3 Profile
20
18
16
14
12
10
8
-1.0
TSS
TES
1.0Kb

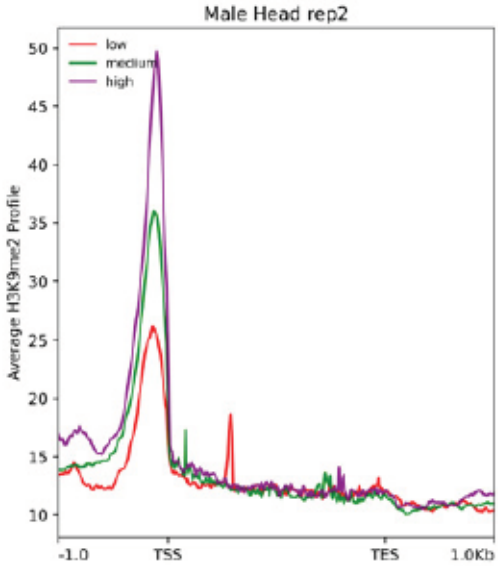

Female Head rep1
low
medium
high
Average H3K9me3 Profile
22
20
18
16
14
12
10
-1.0
TSS
TES
1.0Kb

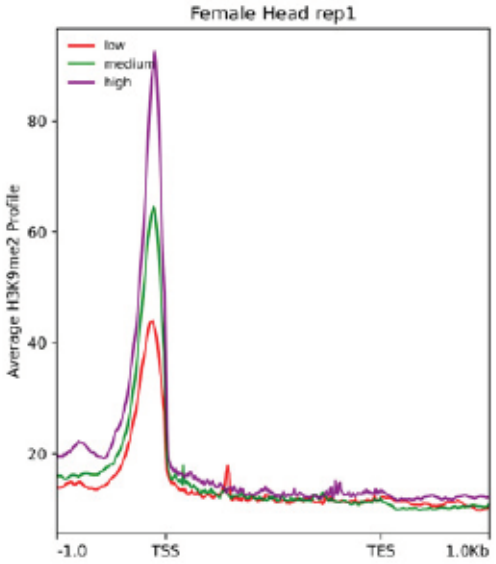

Female Head rep2
low
medium
high
Average H3K9me3 Profile
30
25
20
15
10
-1.0
TSS
TES
1.0Kb

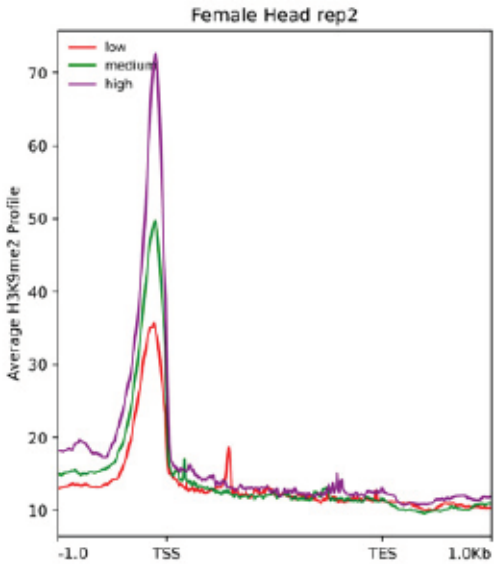


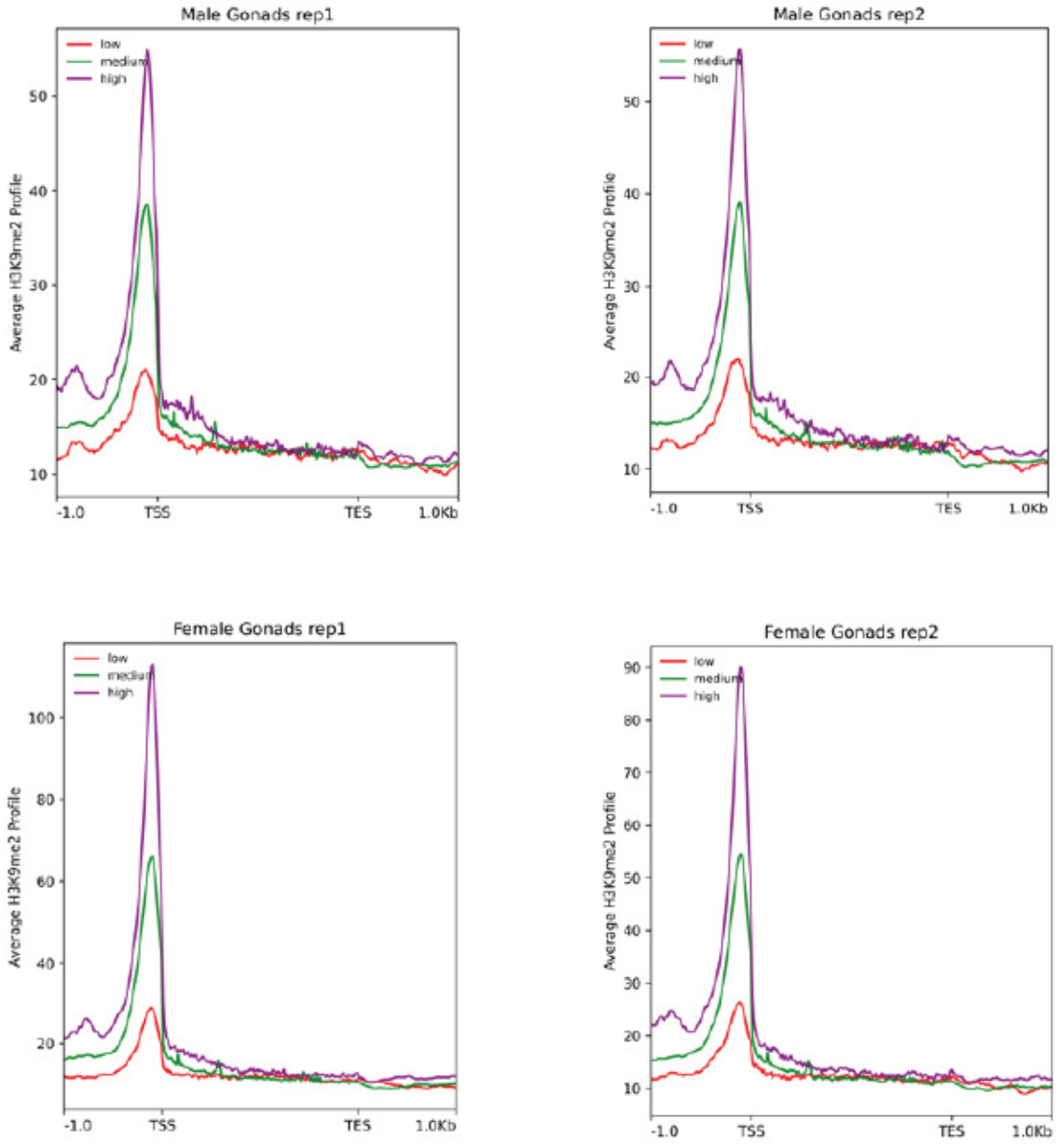
Supplementary Figure 18: Average enrichment of H3K9me3 across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads


Male Head rep1
low
medium
high
Average H3K9me2 Profile
45
40
35
30
25
20
15
10
-1.0
TSS
TES
1.0Kb

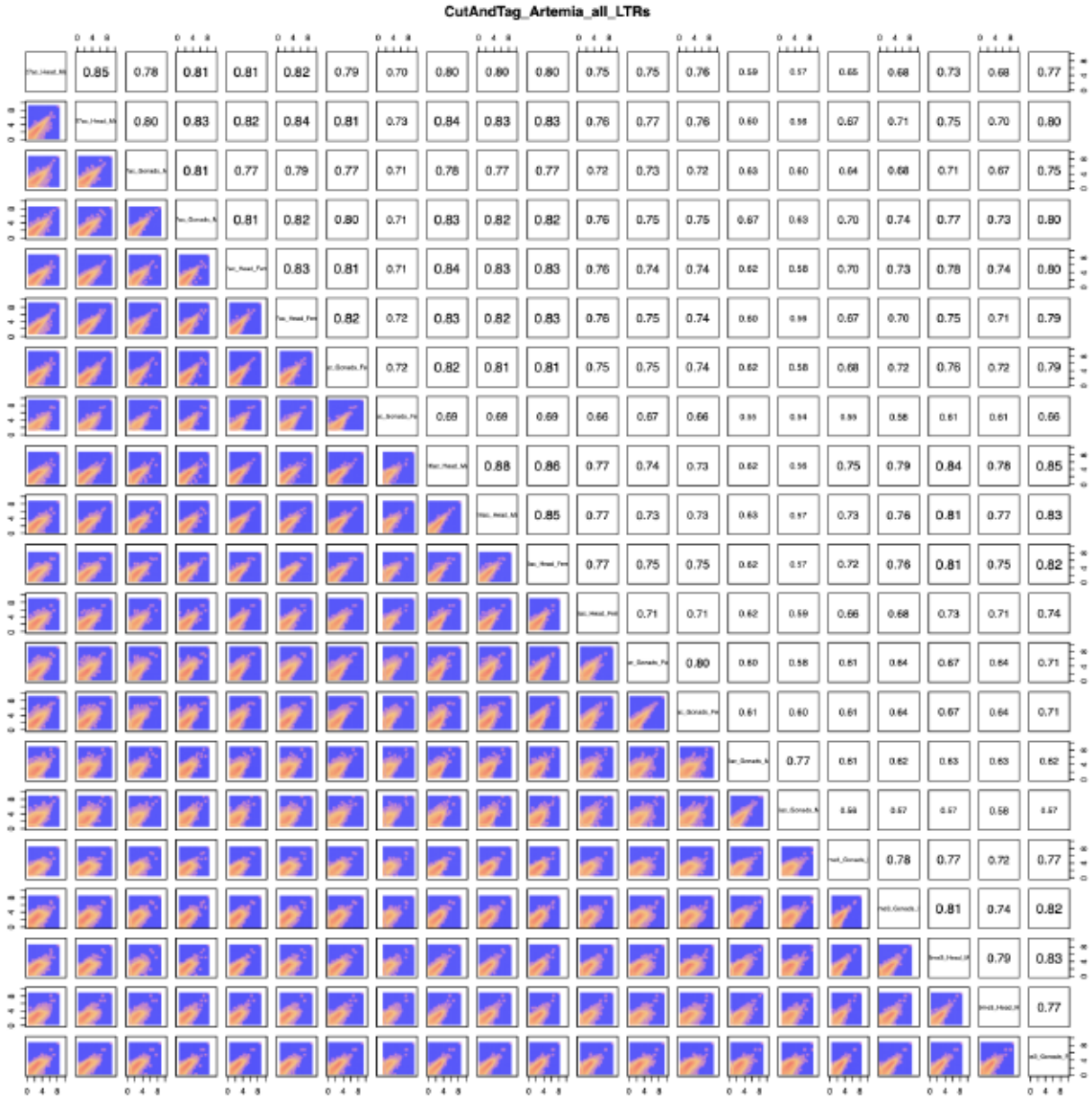

Male Head rep2
low
medium
high
Average H3K9me2 Profile
50
45
40
35
30
25
20
15
10
-1.0
TSS
TES
1.0Kb


Female Head rep1
low
medium
high
Average H3K9me2 Profile
80
60
40
20
-1.0
TSS
TES
1.0Kb

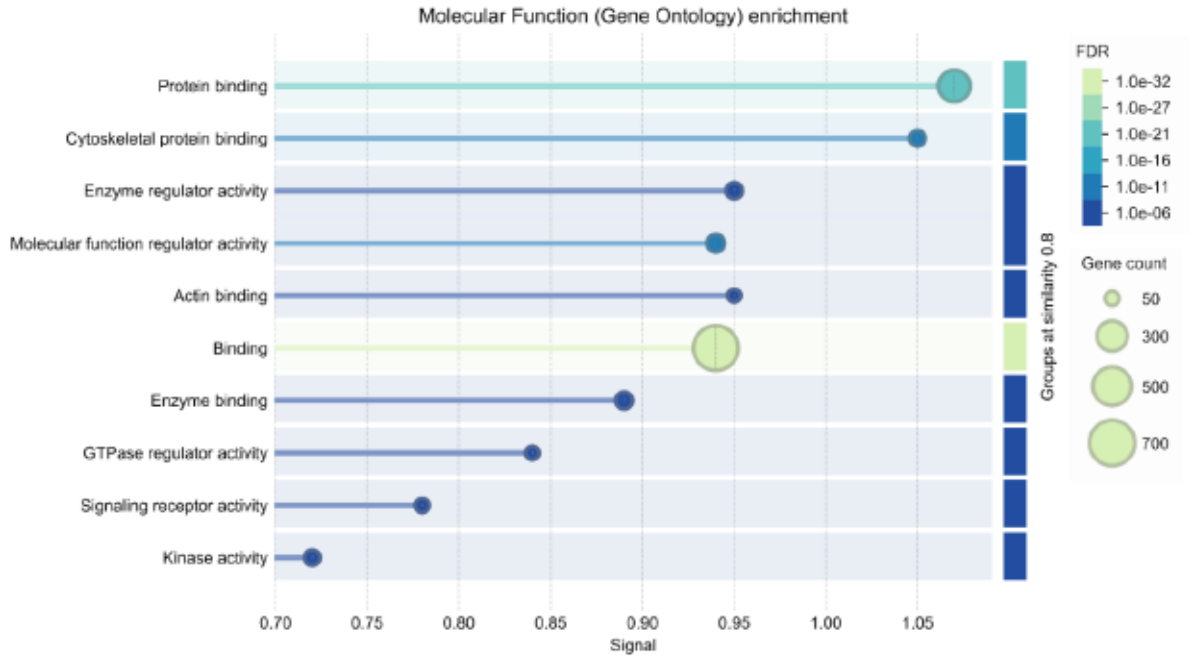

Female Head rep2
low
medium
high
Average H3K9me2 Profile
70
60
50
40
30
20
10
-1.0
TSS
TES
1.0Kb



Supplementary Figure 19: Average enrichment of H3K9me2 across gene length (only autosomal and PAR genes were considered with TPM > 0.5) of different expression levels (low (those less than 30% in TPM across these genes considered), medium (those genes between 30% and 70%) and high (those above 70%)) in both head and gonads



Supplementary Figure 20: Correlation of all samples with all tested histone marks in gonads and heads tissues of both males and females.



Supplementary Figure 21: Functional enrichment of genes with chromatin signature (S12) in gonads

Marks	Tissue	Sex	TEs introns	r-value	z-value	z-observed	*P*-value of z-diff
H3K9me3	Heads	Male	Yes No	0.200 -0.02	0.202 -0.02	14.87	4.89e-50
		Female	Yes No	0.15 -0.03	0.151 -0.03	12.1	1.1e-33
	Gonads	Male	Yes No	0.16 -0.04	0.141 -0.04	12.08	1.29e-33
		Female	Yes No	0.14 -0.14	0.1409 -0.141	18.82	5.07e-79
H3K9me2	Heads	Male	Yes No	0.18 -0.03	0.182 -0.03	14.15	1.71e-45
		Female	Yes No	0.20 0.03	0.203 0.03	11.53	8.9e-31
	Gonads	Male	Yes No	0.20 -0.02	0.2027 -0.02	14.87	4.8e-96
		Female	Yes No	0.22 0.05	0.223 0.05	11.593	4.45e-31

Supplementary Table 1: r-z transformations of coefficients of H3K9me3 and H3K9me2 enrichment and expression those genes with TEs in introns and those without TEs in introns in heads and gonads

Type	Tissue	Autosomes	Pseudo-autosomal (PAR)	Z-linked regions	W-linked genes
Male-biased	Gonads	926	31	22	0
Female-biased	Gonads	594	26	6	7
Male-biased	Heads	10	1	0	0
Female-biased	Heads	10	0	0	5

Supplementary Table 2: sex-biased gene expression in heads and gonads of *A. franciscana* (FDR <0.05, Foldchange >2 & TPM >0.1 for genes in *A. franciscana*)

https://github.com/vkb25/Chromatin-landscape-in-Artemia-franciscana/blob/c8aff7a95879177f6451a347f3dd1c52eee1e2cc/female_biasedchromatintranscriptrep_transcript.xlsx

Supplementary Table 3: GO terms of contrasting chromatin states and female-biased gene expression in gonads of *A. franciscana*

https://github.com/vkb25/Chromatin-landscape-in-Artemia-franciscana/blob/c8aff7a95879177f6451a347f3dd1c52eee1e2cc/male_biasedchromatintranscriptrep_transcript.xlsx

Supplementary Table 4: GO terms of contrasting chromatin states and male-biased gene expression in gonads of *A. franciscana*

https://github.com/vkb25/Chromatin-landscape-in-Artemia-franciscana/blob/f31ed9a097cfeec0c37a5fd27d47cd8de3808388/Gonads_homerResults.html

Supplementary Table 5: Transcription factor binding sites with significant enrichment within H4K16ac CUT&TAG peaks in female gonadal tissues. Analysis was performed using HOMER

https://github.com/vkb25/Chromatin-landscape-in-Artemia-franciscana/blob/f31ed9a097cfeec0c37a5fd27d47cd8de3808388/Somatic_homerResults.html

Supplementary Table 6: Transcription factor binding sites with significant enrichment within H4K16ac CUT&TAG peaks in female head tissues. Analysis was performed using HOMER

https://github.com/vkb25/Chromatin-landscape-in-Artemia-franciscana/blob/d9e9091814fe0611e2ccab2f67a752a25f596f53/supplementary%20files%20on%20mapping%20reads.xlsx

Supplementary Table 7: Total reads mapped, alignment rate, estimated library size and the amount of mapped reads retained for downstream analysis after applying strict filtering steps

Sample ID	Sex	Tissue
SRR8641214	Female	Ovaries
SRR8641213	Female	Ovaries
SRR8641212	Female	Head
SRR8641211	Female	Head
SRR8641218	Male	Testes
SRR8641217	Male	Testes
SRR8641216	Male	Head
SRR8641215	Male	Head

Supplementary Table 8: RNA raw reads together with their SRA numbers used for expression analysis.

Chapter 4: Unravelling Sex Chromosome Evolution and Dosage Compensation Mechanisms in *Cameraria ohridella* (horse-chestnut leaf miner moth)

Vincent Kiplangat Bett[1@], Ariana Macon[1], Marwan Elkrewi[1] and Beatriz Vicoso[1@]

Abstract

Lepidoptera (butterflies and moths) exhibit female heterogamety (ZW system). While the Z chromosome is highly conserved across the clade, the origin, structure, and function of the W chromosome remain elusive. However, the presence and functional role of the W chromosome are not consistent across lineages; some species show no evidence of its involvement in sex determination. Sex chromosome evolution often triggers dosage compensation (DC) mechanisms to balance gene expression between autosomes and sex chromosomes and consequently between sexes. ZW systems were traditionally thought to exhibit only partial DC, but recent studies in Lepidoptera, *Artemia franciscana*, and *Apalone spinifera* suggest a diverse range of compensation strategies, challenging traditional assumptions. While DC often involves chromatin-level regulation, the specific mechanisms in ZW systems are still poorly understood. To explore these questions, we generated a genome assembly of *Cameraria ohridella* (horse-chestnut leaf miner moth, Gracillidae), and combined transcriptomic data from two tissues with CUT&Tag epigenomic profiling targeting both active (H4K16ac, H3K4me3) and repressive (H3K27me3) chromatin marks. This enabled us to characterize the composition and putative acquisition of the W chromosome in this species and to explore the functional contributions of its gene content. Our findings reveal a highly dynamic landscape of sex chromosome evolution in *C. ohridella*, including the ancestral Z (AncZ), a $NeoZ_1$ formed by fusion of AncZ with an autosome, and a $NeoZ_2$ that likely arose via formation of a NeoW. We uncover distinct dosage compensation (DC) patterns across ancestral and neo-sex chromosome regions and also between head and whole body tissues. On the AncZ, DC in head tissues involves repression of gene expression in males (ZZ), and this was consistent with depletion of the active histone mark H4K16ac. In contrast, $NeoZ_1$ regions show upregulation of gene expression in the heterogametic sex (ZW), accompanied by enrichment of H4K16ac. Together, our findings underscore the dynamic nature of sex chromosome evolution and reveal variation in dosage compensation strategies across the Z chromosome, where the $NeoZ_1$ appears to evolve entirely new, Drosophila-like mechanisms rather than co-opting the existing Nematode-like mechanisms of AncZ.

Introduction

Sex chromosomes have independently evolved multiple times in both animals and plants, often as XY (male heterogamety) or ZW (female heterogamety) systems (Graves 2008). They are acquired from pairs of initially homologous autosomes after the emergence of sex-determining factor (s), triggering the accumulation of sexually antagonistic mutations (Wright et al. 2016) which would favor the evolution of recombination suppression between newly formed proto-X and Y (or Z and W) chromosomes. The absence of recombination would result in progressive genetic degeneration of Y chromosome (or W), characterized by accumulation of deleterious mutations, repeat accumulation, and gene loss (Bachtrog 2013). The fusion of sex chromosomes with autosomes to form neo-sex chromosomes was generally considered a rare evolutionary event in systems with female heterogamety (Pokorná et al. 2014) and in particular species with high chromosome number such as Lepidoptera (where ZW females typically have 31 chromosomes) (Anderson et al. 2020). Despite this expectation, emerging evidence points to a surprising number of neo-sex chromosome formations within Lepidoptera (Wright et al.

2024; Mora et al. 2024), including documented case in the leaf miner moth *Cameraria ohridella* (Fraïsse et al. 2017). This raises key questions about the evolutionary origins of these neo-Z chromosomes and whether similar patterns extend across other genera in the Gracillariida. While Z and X chromosomes in winged insects generally have distinct ancestral origins, the Z chromosomes in Trichoptera (Z0) and Lepidoptera (ZW) have deep conservation of gene content (Wright et al. 2024). In contrast, the evolutionary origin and functional significance of the lepidopteran W chromosome remain elusive, largely due to its highly repetitive content and scarcity of protein-coding genes, which makes it challenging to trace its evolutionary relationships with the Z chromosome and across species (Traut et al. 2013). Interestingly, the presence and functional conservation of the W chromosome are not consistent across lepidopteran lineages; for instance, *Samia cynthia* ssp. exhibit no apparent involvement of the W or a neo-W in sex determination and reproduction (Yoshido et al. 2016). In contrast, the W chromosome of the silkworm (*Bombyx mori*) carries a dominant gene, *Fem*, that actively drives the development of female characteristics (Kiuchi et al. 2014). Therefore, deciphering the complete sequences of the W chromosome could also be essential for understanding the genetic mechanisms underlying sex determination, potentially uncovering master regulatory genes. These insights could inform the development of sex-specific pest control strategies, particularly relevant for this clade due to the destructive behavior of its larvae.

The degeneration of the sex-limited chromosome (W or Y) in the heterogametic sex often leads to imbalances in gene copy number, which can disrupt the stoichiometry of protein complexes and reduce organismal fitness and lethality (Veitia 2005). To counteract these effects, many species have evolved sex-specific regulatory strategies referred to as dosage balance and dosage compensation that restore equilibrium in gene expression between sex chromosomes and autosomes, and consequently between the sexes (Lucchesi et al. 2005). These regulatory mechanisms may operate across the entire sex chromosome or target only specific, dosage-sensitive genes, resulting in either chromosome-wide or localized, partial compensation respectively (Gu and Walters 2017). Different organisms have evolved diverse strategies that involve epigenetic regulation to achieve dosage compensation (DC) (Muyle et al. 2021). In XX/XY systems, DC typically involves orchestrated chromosome-wide modulation of gene expression. For instance, in both placental mammals and *C. elegans*, DC is accomplished by repressing transcription in the homogametic (XX) sex (Lau and Csankovszki 2015). In contrast, *Drosophila* achieves DC by upregulating transcription from the single X chromosome in males to match the expression levels of the two X chromosomes in females (Samata and Akhtar 2018). This hyperactivation is driven by the deposition of activating histone modifications, notably H4K16ac, which promotes chromatin decompaction and facilitates increased transcription (Samata and Akhtar 2018). ZW species were initially believed to exhibit only partial dosage compensation, but recent studies in Lepidoptera (moths and butterflies) and Artemia have revealed chromosome-wide compensation mechanisms (Bett et al. 2025; Gu et al. 2019). In *Artemia franciscana*, a well-differentiated region of the Z chromosome in females (ZW) is enriched with the active histone mark H4K16ac, enhancing transcription to balance the two Z copies in males (ZZ) (Bett et al. 2025; Zimmer et al. 2025). In the monarch butterfly, gene expression along the Z chromosome shows a dichotomy of dosage compensation mechanisms: the ancestral Z is depleted of H4K16ac in males, resembling nematode-like compensation, while the neo-Z is enriched with H4K16ac in females, mirroring Drosophila-like compensation (Gu et al. 2019).
In this study, we applied genomic approaches to investigate the evolutionary dynamics and regulatory architecture of complex multiple sex chromosomes in the horse-chestnut leaf miner, *Cameraria ohridella*. We sequenced, assembled, and annotated the genome of *C. ohridella* and

utilizing previously published male and female genomic data, we assess the extent and structure of sex chromosome differentiation. To further explore the functional consequences of this differentiation, we analyze RNA-seq data from head and whole-body tissues, focusing on patterns of dosage compensation across distinct regions of the Z chromosome that were partitioned based on their evolutionary origins. We also examine the influence of sex-biased gene expression on dosage compensation dynamics. Finally, we investigate the chromatin landscape across these Z-linked regions using CUT&Tag profiling of key histone modifications; H4K16ac, H3K4me3, and H3K27me3 to probe their role in mediating dosage compensation. Together, our work sheds light on how genomic, transcriptomic, and epigenetic layers interact during the evolution of complex sex chromosome systems.

Results

Evolutionary Dynamics of Sex Chromosomes

In order to construct a high-quality genome for the female *Cameraria ohridella*, we utilized the PacBio Sequel II SMRT platform, generating 85.7 Gb of long-read PacBio HiFi data (~214-fold coverage). The primary assembly was subjected to four rounds of duplicate purging using consensus reads, culminating in a contig-level assembly. The final contig genome spans approximately 453 Mb, comprising 447 contigs with an N50 of 2.5 Mb and a longest contig of 9.2 Mb (Table S1). Scaffolding using consensus HiFi data resulted in an assembly with an N50 length of ~9.1 Mb, and BUSCO analysis (arthropoda_odb10) indicated a genome assembly completeness of 98.4%, including 92.2% single-copy genes (supplementary figure 1), underscoring the high quality and completeness of the assembly. Previous research has identified the presence of a neo-Z chromosome in *Cameraria ohridella*, formed by the fusion of the ancestral Z chromosome with autosomal one (chromosome 18 in *Bombyx mori*) (Fraïsse et al. 2017). To investigate this further, we computed genome-wide coverage using short-read male and female DNA, applying a window size of 5,000 bp. By analyzing the female-to-male coverage ratio, we identified two superscaffolds, 3 and 30, derived from the differentiated part of the Z chromosome (Fig. 1b). These superscaffolds correspond to the ancestral Z and autosomal chromosome 18 in *Bombyx mori* as previously reported (Fraïsse et al. 2017) and map exclusively to the Z chromosome in *Euspilapteryx auroguttella* (Fig. 1a), a species within the same *Gracillariidae* family as *Cameraria ohridella* (Boyes et al. 2024). This pattern shows that the fusion between the ancestral Z (AncZ) and autosomal chromosome (NeoZ1) occurred prior to the divergence of Gracillariidae subfamilies.

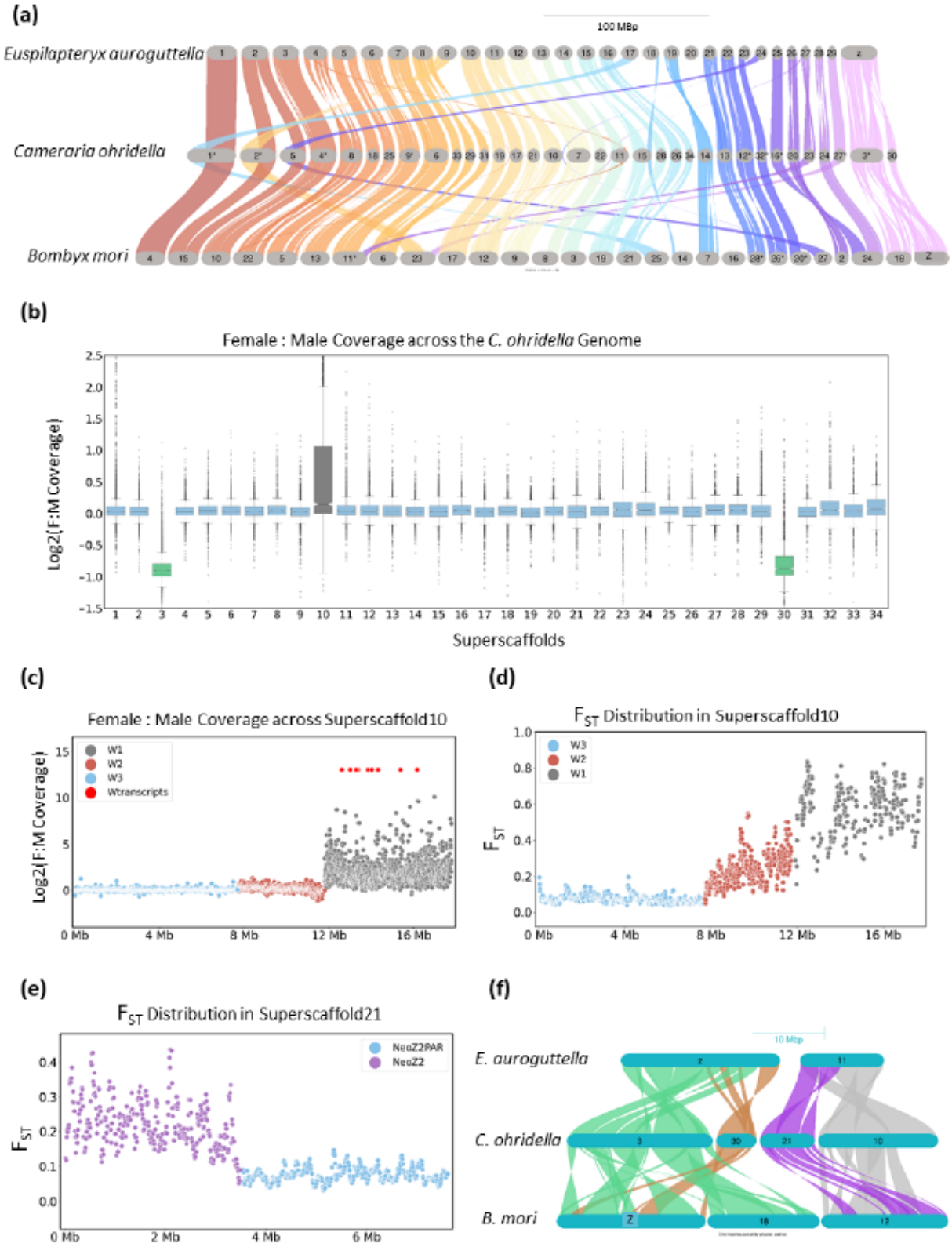


Fig. 1: **Identification and extent of differentiation of Z-linked and putative W-linked chromosomes** (a) Synteny across *Bombyx mori*, *Cameraria ohridella* and *Euspilapteryx auroguttella* (species within the Gracillariidae family) estimated using GENESPACE. (b) Female-to-male whole genome sequencing coverage over a 5kb window across the *Cameraria*

ohridella genome assembly. (c) Female-to-male whole-genome sequencing coverage over a 5 kb window, alongside putative W-linked transcripts in superscaffold 10. (d) Female-to-male F_{ST} distribution over 10kb windows across superscaffold 10. (e) Female-to-male F_{ST} distribution over 10kb windows across superscaffold 21. (f) Synteny analysis of selected superscaffolds (sex chromosomes) of *Cameraria ohridella*, with distant related species *Bombyx mori* and closely related species *Euspilapteryx aurogutella* (same family as *C. ohridella*).

Superscaffold 10 had a strongly increased female:male coverage ratio compared to other chromosomes, suggesting it may correspond to the W chromosome. The presence of genetic variants linked to the W chromosome are often used to detect regions of ZW that recently stopped recombining, therefore, we used pooled DNA of female and male to quantify the genetic divergence (fixation index F_{ST}) between them. Integrating both coverage and F_{ST} analyses, we identified three distinct evolutionary strata within the W chromosome (Figs. 1c and 1d). The first stratum (W1), spanning approximately 12Mb to 17Mb, is characterised by increased female coverage and elevated F_{ST} values, indicative of an older, highly differentiated W-linked region (Figs. 1c and 1d). The second stratum (W2), extending from 8Mb to 12Mb, shows a moderate increase in female coverage combined with pronounced F_{ST}, reflecting a more recent non-recombining region of the W chromosome. In contrast, the segment from 0Mb to 8Mb exhibits similar coverage and F_{ST} values in both sexes, which likely defines a pseudoautosomal region (W3) where recombination persists (Figs. 1c and 1d). To further validate the W chromosome, we employed a K-mer approach to identify and assemble female-specific transcripts, which are expected to correspond to W-linked sequences (see Methods). To assess whether these transcripts are W-linked, we examined their homology across the genome. The majority mapped to superscaffold 10 (supplementary fig. 2), specifically aligning to the W1 region (Fig. 1c), further supporting its identity. To further confirm the female specificity of the putative W chromosome, we selected seven fragment sequences from the (W1) for verification using polymerase chain reaction (PCR) in ten individuals (five females and five males). Of these, six consistently produced female-specific bands, reinforcing their association with the W chromosome (supplementary fig. 3). Our characterization of the putative W chromosome reveals its distinct evolutionary trajectory. For instance, it has accumulated a remarkably high repeat content (79.42%), significantly exceeding the female genome's overall repeat content (56.84%) (supplementary fig. 4). This accumulation is accompanied by a low gene density and reduced gene expression (~56% of genes in W1 have TPM <1), with a median TPM of ~0.4 substantially lower than the autosomal median of ~4 TPM. The W1 of the putative W chromosome maintains a slightly elevated GC content (38.32%) compared to the genome-wide average of 38.08% (supplementary fig. 5).
Sex chromosomes often evolved from ordinary pairs of homologous autosomes. To find chromosomes linked to the putative W chromosome (superscaffold 10), we examined the distribution of reciprocal best hits of genes from superscaffold 10 across the *C. ohridella* genome. This analysis revealed that most genes from the W2 and W3 regions of superscaffold 10 corresponded to superscaffold 21 (supplementary fig. 6). Furthermore, synteny analysis showed that both superscaffold 10 (W) and superscaffold 21 ($NeoZ_2$) share homology with *B. mori* chromosome 12 and *E. auroguttella* chromosome 11 (Fig. 1f), suggesting that superscaffold 21 is a second neo-Z chromosome, which we refer to as $NeoZ_2$. Despite the similarity in female-to-male coverage across superscaffold 21 and also with autosomal chromosomes (supplementary figs. 7), F_{ST} analysis of superscaffold 21 revealed two strata: a region with elevated F_{ST} between males and females, consistent with a non-recombining ZW region, and a distal region with uniform F_{ST}, indicating ongoing recombination and possibly representing a pseudoautosomal segment of $NeoZ_2$ (Fig. 1e).

To explore the potential origins of genes in the W1 region of *C. ohridella* superscaffold 10, we conducted a reciprocal best hits analysis against the rest of the genes in the genome. Our results showed that the W1 genes were not enriched on any particular chromosome but rather distributed across the genome (Fig. 2a), suggesting that these genes were likely translocated from multiple genomic regions into W1 or have underwent genetic differentiation (from $NeoZ_2$ homology). Moreover, we observed gene expansion within the W1 region through gene duplications (Fig. 2a, supplementary figs. 8). Since transposon-encoded genes are typically highly repetitive and mobile within the genome, we investigated their association with the W1 genes. We found that 33 out of 106 W1 genes contained TEs, however, among the 41 duplicated genes within the W1 region, only 6 had TEs. Further analysis of the W1 genes' reciprocal best hits across the *Bombyx mori* genome revealed orthologs on multiple chromosomes, including some present on the W chromosome (Fig. 2b).

The genetic differentiation of the ancestral Z chromosome (AncZ) in *C. ohridella* identified regions with elevated F_{ST} values (Figs. 2c and 2d), a pattern consistent with the presence of corresponding W-linked sequences. Notably, a reciprocal best-hit analysis comparing all genes against each other revealed that these high-F_{ST} regions are homologous to each other (Fig. 2f). Further investigation, focusing on genes located outside of the primary superscaffolds (specifically scaffolds 35 to 220) for putative W scaffolds with genes corresponding to these areas, identified scaffold 48 having pronounced levels of female-to-male coverage (Fig. 2e). Additionally, scaffold 43 contains a specific region, spanning from 1.77 Mb to 3.4 Mb, characterized by elevated F_{ST} levels (supplementary fig 9), suggesting a possible link to sex-specific genomic differentiation.

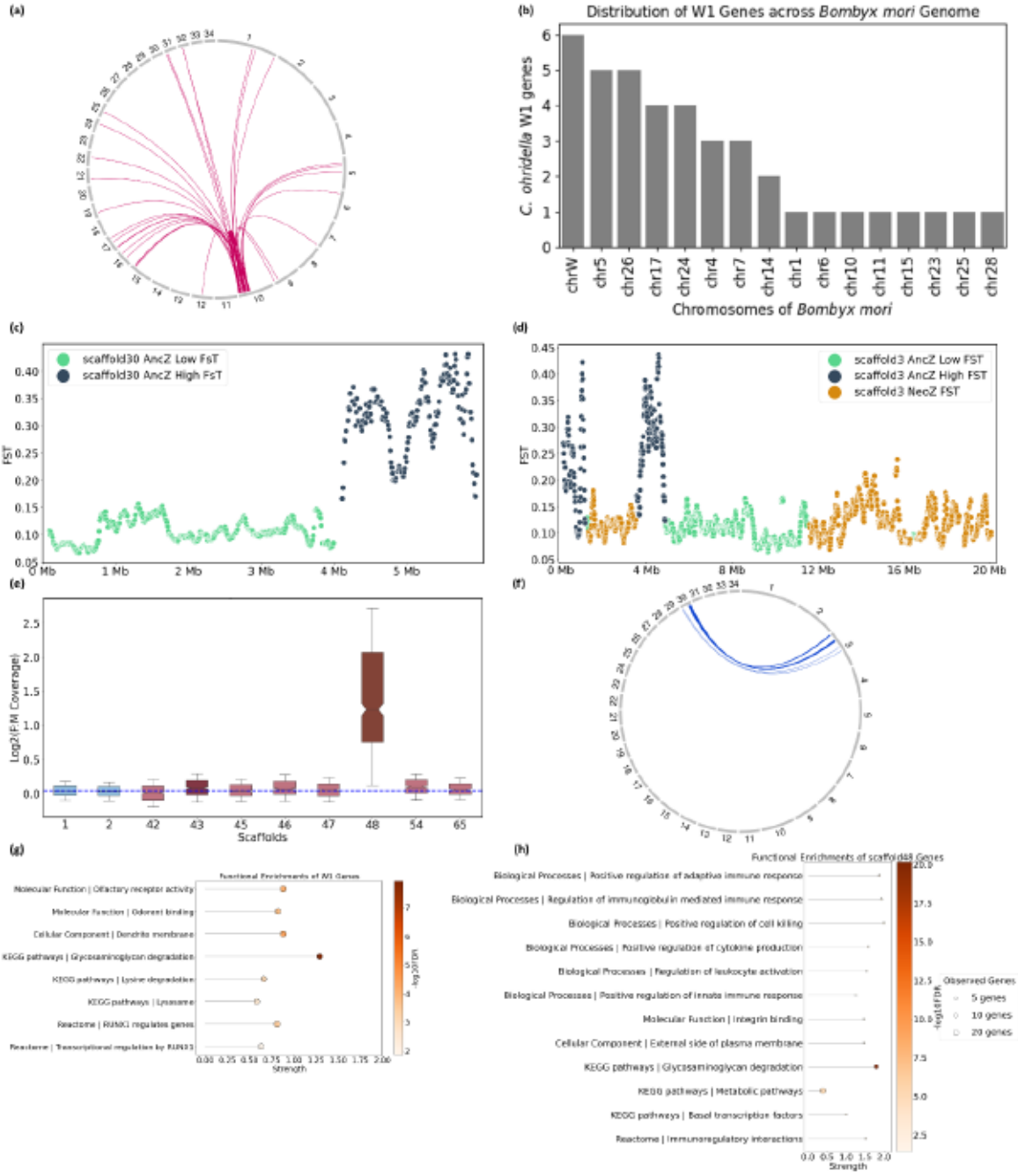


Fig 2: **Dynamics of sex chromosome evolution and functional enrichment of putative W-linked genes** (a). Homology of genes in first stratum (S0) in superscaffold 10 identified through an all-vs-all BLAST analysis, filtering for reciprocal best hits with a minimum sequence identity of 50 bp. (b). Distribution of *C. ohridella* W1 gene orthologs across the chromosomes of *Bombyx mori*. (c). F_{ST} distribution across AncZ and $NeoZ_1$ of the superscaffolds 3 (d). F_{ST} distribution across anc-Z of the superscaffolds 30 (e) Female-to-male coverage of putative W scaffolds selected based on gene homology with those in high-FST regions of AncZ. (f). High-F_{ST} regions of AncZ identified as homologous to one another through reciprocal best-hit analysis. (g) GO terms of genes in the W1 region of the superscaffold 10 (h) GO terms of genes in the scaffold 48, the putative W chromosome, with some genes having homology with genes in High-F_{ST} regions of AncZ.

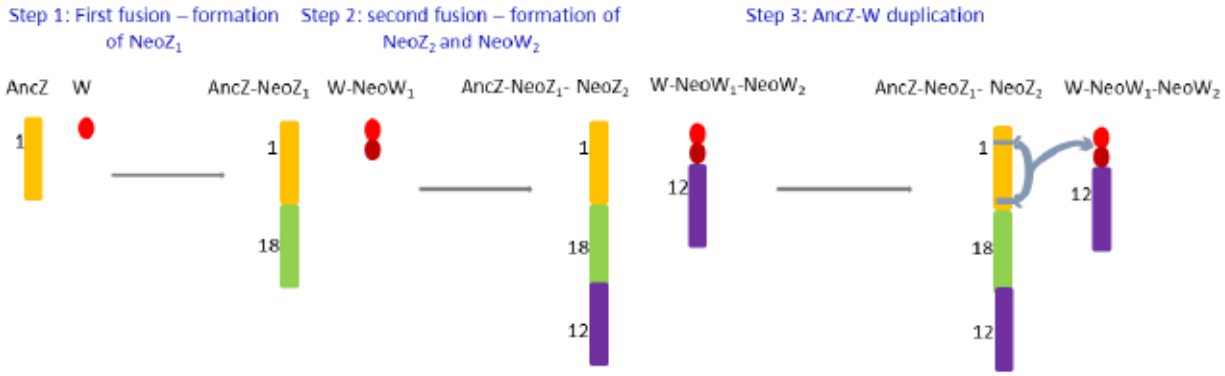


Fig 3: **Proposed evolutionary dynamics of sex chromosome evolution in *Cameraria ohridella*.** This model proposes an evolutionary dynamic of sex chromosome evolution in *Cameraria ohridella* species with ZW system. Initially, there exists an ancestral Z (corresponding to chr1 in *B. mori*) and the W chromosome which is small and likely heterochromatic. Then, AncZ likely fused with an autosome (corresponding to chr18 in *B. mori*) resulting in a neo-sex chromosome (AncZ-NeoZ1) and its homologous pair ($NeoW_1$) underwent rapid degeneration. The second fusion event occurred that involved W-$NeoW_1$ with an autosomal chromosome (corresponding to chr12 in *B. mori*) creating a W-$NeoW_1$-$NeoW_2$. Concurrently, the Z chromosome acquired a homologous pair of chromosome 12, forming a chromosome AncZ-NeoZ1-$NeoZ_2$. The final step is likely to involve AncZ-W duplication and structural rearrangements such as degeneration, translocation and gene duplications in the W. In summary, sex chromosome system consists of a large, multi-origin Z composed of ancestral and neo-sex chromosome regions (1, 18, and 12 corresponding chromosomes in *B. mori*), and a degenerate W chromosome containing ancestral W and portions of chromosome 12.

Dosage compensation in sex chromosomes with dynamic evolution

Degeneration of the W chromosome and the concomitant gene loss impose dosage constraints on the heterogametic sex, because transcriptional output generally correlate with gene copy number. The haploidy of the Z chromosome disrupts regulatory equilibrium between sex-linked and autosomal loci. To mitigate these imbalances, Z-linked dosage compensation mechanisms evolve to modulate transcriptional output. As a secondary effect, these mechanisms also equalize sex-linked gene expression between males and females (dosage balance). We therefore examined dosage balance and compensation across the AncZ, $NeoZ_1$, and $NeoZ_2$ and autosomal chromosomal regions by analyzing sex-specific gene expression in head and whole-body tissues. In somatic tissue (heads), $NeoZ_1$ genes exhibited expression levels comparable to autosomes and showed no significant differences between males and females, suggesting complete dosage balance and compensation (P-value < 0.1, Fig. 4a). In contrast, $NeoZ_2$ showed reduced expression in females relative to males, indicating incomplete dosage balance (P-value=0.0003, Fig. 4c). The AncZ region presented a more complex pattern: although overall expression was higher in females than in males (P-value=0.0001), this bias was largely driven by genes in high F_{ST} regions (P-value=1.9e-06, Fig. 4c) which also point out that the female-bias in expression is likely simply due to the excessive copies of these genes on the W chromosome. However, the genes in low F_{ST} regions exhibited balanced expression between sexes (P-value=0.14, Fig. 4c). Nevertheless, genes in both high and low F_{ST} regions of the AncZ showed significantly lower expression compared to autosomes in both sexes (P-value < 0.01,Fig. 4a), pointing to incomplete dosage compensation. These findings highlight

the dynamic nature of sex chromosome evolution and suggest that dosage compensation mechanisms may evolve in a region-specific manner along the Z chromosome.

In whole-body tissues, regions of the ancestral Z (AncZ) chromosome with low F_{ST} as well as the $NeoZ_1$ chromosome that previously showed complete dosage balance between sexes in somatic tissue (heads) exhibited a different expression pattern: gene expression was significantly reduced in females compared to males (Fig. 4d). Similarly, high F_{ST} regions of the AncZ, which were previously overexpressed in females, also showed reduced expression in females relative to males in whole-body samples (Fig. 4d). While expression of $NeoZ_1$ in males remained comparable to autosomes ($NeoZ_1$:A ratio = 0.93, *P*-value = 0.35, Fig. 4b), females showed a significant reduction in $NeoZ_1$ expression relative to autosomes ($NeoZ_1$:A ratio = 0.583, $P = 0.0004$, Fig. 4b), indicating incomplete dosage compensation. Both the low F_{ST} regions of the AncZ and $NeoZ_1$ were enriched for male-biased genes, suggesting that gonadal signals may influence expression patterns in whole-body tissues (supplementary fig 10).

Importantly, when sex-biased genes (those with ≥2-fold expression differences between sexes and adjusted *P*-values < 0.05) were excluded, the expression differences between males and females in $NeoZ_2$ and high F_{ST} were no longer statistically significant (supplementary fig 11). However, the reduced expression of $NeoZ_1$ and the low F_{ST} regions of AncZ in females persisted (supplementary fig 11). These findings highlight the role of tissue context and sex-biased gene content in shaping dosage regulation across sex chromosomes.

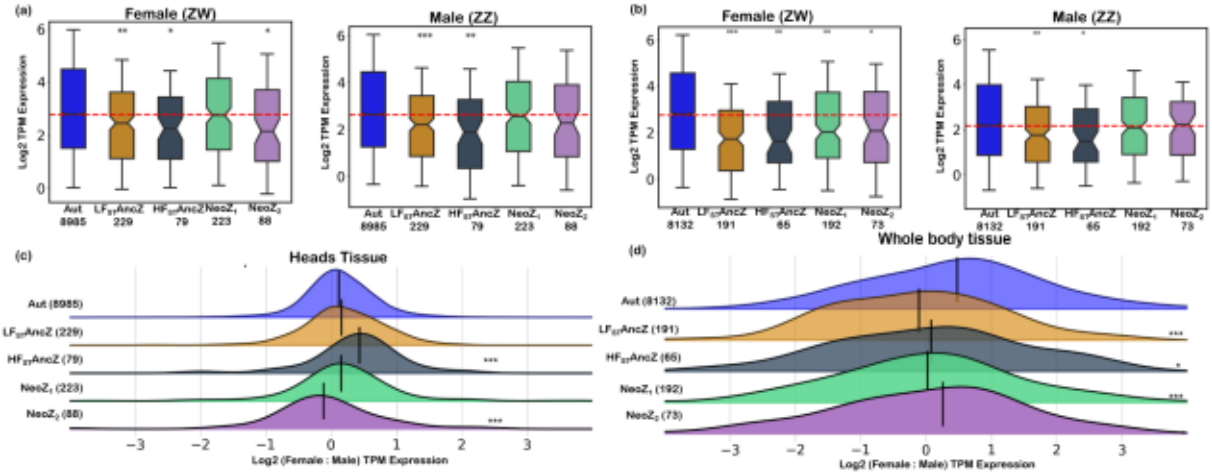


Fig 4: Patterns of dosage compensation between the Z-linked chromosomes and autosomes in heads and whole body tissues of both males and females. Distributions of Log_2 transformed TPM expression values are given for two regions of ancestral Z (chr1 in *B. mori*), $NeoZ_1$ (chr18 in *B. mori*), neo-Z_2 (chr12 in *B. mori*) and autosomes for each sex and two tissues in Cameraria ohridella. The red dotted lines in box plots represent the overall expression median in each sex. The vertical black lines represent median values. Statistical significance of Wilcoxon signed-rank test applied to each data set between autosomes and each Z-linked region and indicated by stars (*P-value < 0.05, ** is P-value < 0.005, and *** is P-value < 0.0005).

Dual Strategies, One Chromosome: dosage compensation mechanisms on the Z

Previous studies of both model organisms such as *Drosophila*, mammals, and the worm *Caenorhabditis elegans*, and non-model species like the brine shrimp *Artemia franciscana* and the green anole suggest that dosage compensation is often mediated through epigenetic regulation of gene expression. These regulatory processes modulate gene expression by altering chromatin structure, often through the targeted addition of DNA methylation or repressive histone modifications to reduce transcription in the homogametic sex, or through the enrichment of activating histone marks that upregulate expression in the heterogametic sex. In Lepidoptera, a recent study on monarch butterflies revealed distinct dosage compensation strategies on the ancestral Z and neoZ chromosomes (Gu et al. 2019). The ancestralZ showed transcriptional downregulation in the homogametic sex (males), associated with a depletion of active histone mark, while the neoZ exhibits transcriptional upregulation in the heterogametic sex (females), marked by an enrichment of the activating histone modification H4K16ac (Gu et al. 2019). These findings highlight the diverse and region-specific epigenetic mechanisms through which dosage compensation evolves across taxa. In order to explore the chromatin landscape underlying dosage compensation along the Z chromosome in *Cameraria ohridella*, we examined the distribution of key histone modifications associated with transcriptional regulation. Specifically, we focused on H3K4me3, a mark of active promoters; and H4K16ac, an activating modification associated with transcriptional hyperactivation in the heterogametic sex in species such as *Drosophila*, *Artemia franciscana*, and the green anole. We also examined the repressive modification H3K27me3; however, unlike H3K4me3 and H4K16ac, which showed clear enrichment patterns correlated with gene expression, H3K27me3 enrichment did not exhibit a clear relationship with expression levels (supplementary fig 12) and therefore no further downstream analysis were conducted in this histone modification

To assess the epigenetic contribution to dosage compensation, we examined the distribution of histone modifications (using uniquely mapped reads) across genes (including flanking ±500 bp regions) on the AncZ, NeoZ$_1$, and autosomes in both sexes. In males, both Z-derived regions exhibited a significant reduction in H3K4me3 enrichment relative to autosomes (Fig. 5a). In contrast, females displayed comparable levels of enrichment across autosomes, AncZ, and NeoZ$_1$ regions (Fig. 5a). The male-specific reduction was most evident around transcription start sites (TSS), with enrichment levels on the AncZ, NeoZ$_1$, and autosomes becoming similar toward transcriptional end sites (TES) (supplementary fig 13). In contrast, H4K16ac exhibited a pronounced sex- and region-specific distribution. In females, this modification was markedly enriched on the NeoZ$_1$, whereas the AncZ showed levels similar to those of autosomes (Fig. 5b). In males, the pattern was reversed: the AncZ was significantly depleted for H4K16ac, while the NeoZ$_1$ retained autosome-like levels (Fig. 5b).

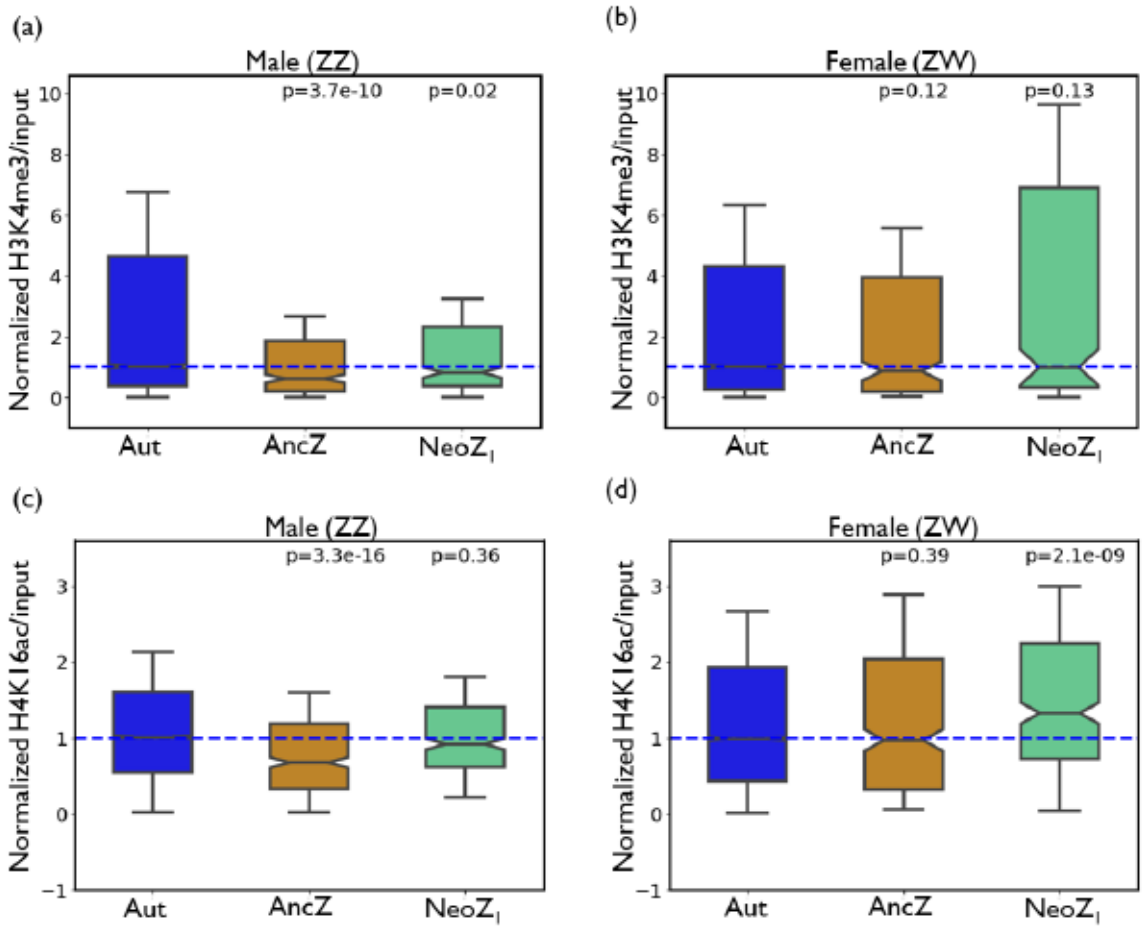


Fig 5: Epigenetic landscape between the Z chromosome and autosome in head tissues of males and females of *Cameraria ohridella*.

Discussion

Here, we uncover the evolutionary trajectory of a complex multi-sex chromosome system in *Cameraria ohridella*. Our analysis reveals that the presence of ancestral Z chromosome and a neo-sex chromosome ($NeoZ_1$), corresponding to the Z chromosome and chromosome 18 in *Bombyx mori*, and both mapped exclusively to the Z chromosome in *Euspilapteryx auroguttella*, a closely related species within the *Gracillariidae* family. This pattern shows that the fusion of the ancestral Z with an autosome likely occurred prior to the divergence of the two subfamilies or may even predate the diversification of *Gracillariidae* as a whole. Fusions between sex chromosomes and autosomes are unexpectedly widespread in Lepidoptera, even though, under a null model where all chromosomes have an equal probability of fusion, such events should be rare in clades with high chromosome numbers such as Lepidoptera ($n = 31$) (Anderson et al. 2020). In addition to the fusion between the AncZ and $NeoZ_1$, our analysis reveals the involvement of another autosomal synteny block corresponding to chromosome 12 in *Bombyx mori*. We favor the hypothesis that a W-autosome fusion occurred in this lineage, resulting in a system with two distinct Z chromosomes. This configuration may resemble the multiple sex chromosome systems described in *Danaus chrysippus* (Martin et al., 2020; Smith et al., 2016) and *Ivela similis* (Mora et al., 2024).

Repetitive sequences accumulating in superscaffold 10 agree with previous CGH studies (Dalíková et al. 2017), which suggested that the W chromosome of *C. ohridella* has accumulated repeats. Around two-thirds is heterochromatic, while the remaining euchromatic part likely represents a younger stratum formed by fusion of the ancestral W with an autosome (Dalíková et al. 2017). Shared genes on the W and ancestral Z chromosomes are unlikely to represent ancient relics; instead, they may result from later duplications of Z-linked genes onto the W chromosome. Comparable Z-W duplications have been documented in the peppered moth (*Biston betularia*) (Hof et al. 2013), where they hypothesized that such transpositions may originate and persist through ectopic recombination between subtelomeric regions (Hof et al. 2013).

The expression patterns of the three Z chromosome regions in *Cameraria ohridella* reflect distinct dosage compensation dynamics. Reduced ancestral Z expression (both highF_{ST} and lowF_{ST}) in each tissue in both sexes relative to autosomes mirrors trends in other Lepidoptera and may result from male-specific down-regulation of the Z (Walters and Hardcastle 2011; Huylmans et al. 2017; Walters et al. 2015), as demonstrated in *B. mori* embryos where *Masc* knockdown led to increased expression of Z-linked genes (Kiuchi et al. 2014). In whole-body samples, the expression pattern of the $NeoZ_1$ chromosome differed markedly from that observed in somatic tissues, with females showing substantially lower expression than males. This disparity may reflect the influence of the gonads, which often constitute a considerable proportion of body mass in invertebrates. Several mechanisms could underlie this pattern, including the absence of dosage compensation, the occurrence of meiotic sex chromosome inactivation (or general down-regulation) on the Z chromosome, the preferential accumulation of testis-specific genes, or a combination of these factors (Huylmans et al. 2017). Removing sex-biased genes had little impact on the overall pattern: the statistical significance remained unchanged and only a minor reduction (≈2%) in mean expression relative to the autosomes was observed. This contrasts with observations in *B. mori*, where exclusion of such genes decreased the difference in mean absolute male to female expression ratio by approximately 25% in gonads (Walters and Hardcastle 2011). The limited effect in *C. ohridella* may reflect developmental stage, population variation, or cell-type composition, factors that likely underlie differences in compensation between head and whole-body tissues. Dissecting these influences will require analyses across specific cells, various tissues, and developmental stages, that may be facilitated by single-cell RNA sequencing. Testing multiple thresholds is also advisable to evaluate robustness. Percentile-based cutoffs (e.g., the top 80% of expressed X-linked genes) may provide a more rigorous alternative to fixed thresholds, though the present analysis considers all genes.

We observe enrichment of H4K16ac on the $NeoZ_1$ in females, whereas this mark is depleted on the AncZ in males, indicating independent evolution of distinct dosage compensation mechanisms across the two Z segments. This mirrors the pattern seen in *D. plexippus*, despite it involves the fusion of a completely different autosome (chromosome 16) with the ancestral Z (supplementary fig. 14). This is at odds with the well-studied situation of Drosophila neo-sex chromosomes, which largely appear to progressively co-opt the ancestral mechanism of dosage compensation for their regulation. Over 30 Z-autosome fusions have now been detected in Lepidoptera (Wright et al. 2024) with many more likely to follow with the generation of more chromosome-level genome assemblies. An in-depth investigation of their transcriptional and chromatin profiles may in the future shed light on why these fusions arose in the first place, on whether dual mechanisms of dosage compensation are the norm throughout the clade, and on why this may be.

Materials and Methods

DNA Extraction and Sequencing

We collected 30 *Cameraria ohridella* females from European horse-chestnut trees in Vienna, Austria for high molecular weight DNA extraction using the Qiagen Genomic-Tip 20/G Kit. The DNA was sequenced on a PacBio Sequel II SMRT cell at the Vienna Biocenter sequencing facility.

Genome Assembly and its Quality Assessment

The consensus sequences from the reads were assembled using Hifiasm with default parameters. The primary assembly was then processed to remove duplicates with consensus reads, using purged_dups over four iterations. To scaffold the purged assembly, we employed LongStitch (ntlink + arcs). We then assessed the completeness of the genome assembly using BUSCO v5.2.2 with the arthropoda_odb10 database (n = 1,013 single-copy orthologs). To further evaluate genome completeness based on k-mer representation, we used Merqury v1.4, which compares the assembled genome against k-mers derived from Illumina short reads. K-mer databases were generated using Meryl v1.4, and completeness evaluation was performed in Merqury using default settings, providing insights into the representation of the nuclear genome.

Coverage and F_{ST} Analysis

Male and female Illumina short-read genomic data, as described by Fraïsse et. al 2017 (Fraïsse et al. 2017), were aligned to the reference genome using Bowtie2 (v2.4.5) with the --end-to-end and --sensitive parameters. To retain only uniquely mapped reads, SAM files were filtered to exclude alignments containing the "XS" tag, which indicates secondary alignments, using grep -vw "XS". Genome-wide coverage was then computed in non-overlapping 5 kb windows using the soap.coverage tool. To investigate population differentiation between sexes, we used aligned sequencing reads from two pooled DNA libraries; each representing 10 male and 10 female individuals. Alignments with a mapping quality (MAPQ) score below 20 were excluded. The remaining alignments were subsequently sorted using the view and sort functions of samtools v1.18. We then generated a pileup file of the filtered male and female alignments using the mpileup function in samtools. To quantify genetic differentiation, we employed the Popoolation2 pipeline. A synchronized (sync) file was produced using the sync module of Grenedalf v0.2.0 and used as input for Popoolation2's fst-sliding.pl script. We estimated F_{ST} values between male and female pools across the genome using non-overlapping sliding windows of 10 kb. The analysis was conducted with the following parameters: --suppress-noninformative --min-count 5 --min-coverage 20 --max-coverage 2000 --min-covered-fraction 0.0 --window-size 10000 --step-size 10000 --pool-size 20

Repeat Annotation and Gene Prediction

A *de novo* consensus repeat library was generated using RepeatModeler, and repetitive elements were annotated across the genome with RepeatMasker using this library. To quantify repeat content, the genome was divided into non-overlapping 10,000 bp windows, and a custom Python script was used to calculate the proportion of repetitive elements from the RepeatMasker output. GC content was estimated for the same window size using the '-pattern GC' option in bedtools nuc. Gene prediction was performed on the soft-masked genome using BRAKER v3.0.8, incorporating RNA-seq data from head and whole-body tissues, along with protein sequences from OrthoDB v12 (Arthropoda). The resulting gene models were refined using AGAT to remove transcript isoforms, incomplete genes lacking start and/or stop codons,

and open reading frames shorter than 100 bp. Protein and coding sequences were extracted with the getAnnoFastaFromJoingenes.py script. The completeness of the annotated gene set was assessed using BUSCO v5.2.2.

RNA-seq Processing and Differential Expression Analysis

Cameraria ohridella specimens that were collected from European horse-chestnut trees in Vienna, Austria, between May and September 2024, were sexed, identified and stored at -80 °C until their subsequent processing. RNA was extracted from adult male and female specimens, with each biological replicate comprising either pooled heads (five individuals) or a single whole-body sample. RNA-seq libraries were prepared and subjected to high-throughput sequencing. Raw reads were pseudoaligned to the annotated coding sequences (CDSs) of *C. ohridella* using Kallisto (v 0.50.1) with default parameters to estimate transcript abundances, reported as transcripts per million (TPM) and estimated counts. Differential gene expression (DGE) analysis was performed in R using the DESeq2 package (v4.3.2). A likelihood ratio test (LRT) framework was employed to assess sex-biased transcriptional differences within each tissue. Genes were considered significantly differentially expressed if they exhibited a log2 fold change >1 (i.e., >2-fold change) and a Benjamini-Hochberg-adjusted false discovery rate (FDR) <0.05. Differentially expressed genes were further categorized as male-biased or female-biased based on the direction of expression change.

CUT&Tag Library Preparation and Processing

For each replicate, head tissues from four field-collected adult individuals were pooled and permeabilized for 30 minutes in an ice-cold solution containing 10 mg collagenase in 1 ml Enzyme Dissociation Buffer. CUT&Tag libraries were then prepared using the Active Motif Kit, following the protocol described in Chapter 3. We also checked for the quality for the paired end reads using fastqc. Overrepresented sequences were filtered out using the bbduk.sh package of bbmap tool (Bushnell 2014) with the options "*ref = overrepresented.fasta k=25 hdist=1*." Sequencing adapters were then trimmed with Trim Galore under parameters "*--paired --nextera -q 30 --length 20*". The trimmed paired-end reads were aligned to *C. ohridella* reference genome using bowtie2 v2.4.5 (Langmead and Salzberg 2012) with these options: "*--local --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700.*" The resulting alignments were filtered to remove secondary alignments with the options "grep -v 'XS:i:'". To generate average plots of chromatin signals around genes for each replicate, we used deepTools/bamCompare to calculate normalized coverage in 20-bp bins "*bamCompare -b1 sample.bam -b2 input.bam --normalizeUsing RPKM --scaleFactorsMethod None --binSize 20 --operation ratio --minMappingQuality 30 --minFragmentLength 50*". The resulting files were processed with deepTools/computeMatrix, and the matrices were visualized using the plotProfile function.

References

Anderson, Nathan W., Carl E. Hjelmen, and Heath Blackmon. 2020. "The Probability of Fusions Joining Sex Chromosomes and Autosomes." *Biology Letters* (England) 16 (11): 20200648. https://doi.org/10.1098/rsbl.2020.0648.

Bachtrog, Doris. 2013. "Y-Chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration." *Nature Reviews Genetics* 14 (2): 113–24. https://doi.org/10.1038/nrg3366.

Bett, Vincent Kiplangat, Minerva Susana Trejo-Arellano, and Beatriz Vicoso. 2025.

"Chromatin Landscape Is Associated With Sex-Biased Expression and Drosophila-Like Dosage Compensation of the Z Chromosome in Artemia Franciscana." *Molecular Biology and Evolution* 42 (5): msaf085. https://doi.org/10.1093/molbev/msaf085.

Boyes, Douglas, David C. Lees, Ian Sims, Dominic Phillips, and Clare Boyes. 2024. "The Genome Sequence of the Yellow-Dotted Stilt, Euspilapteryx Auroguttella Stephens, 1835." *Wellcome Open Research* (England) 9: 226. https://doi.org/10.12688/wellcomeopenres.21284.1.

Dalíková, Martina, Magda Zrzavá, Irena Hladová, et al. 2017. "New Insights into the Evolution of the W Chromosome in Lepidoptera." *Journal of Heredity* 108 (7): 709–19. https://doi.org/10.1093/jhered/esx063.

Fraïsse, Christelle, Marion A. L. Picard, and Beatriz Vicoso. 2017. "The Deep Conservation of the Lepidoptera Z Chromosome Suggests a Non-Canonical Origin of the W." *Nature Communications* 8 (1): 1486. https://doi.org/10.1038/s41467-017-01663-5.

Graves, Jennifer A. Marshall. 2008. "Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes." In *Annual Review of Genetics*, vol. 42. no. Volume 42, 2008. Annual Reviews. https://doi.org/10.1146/annurev.genet.42.110807.091714.

Gu, Liuqi, Patrick F. Reilly, James J. Lewis, Robert D. Reed, Peter Andolfatto, and James R. Walters. 2019. "Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly." *Current Biology* 29 (23): 4071-4077.e3. https://doi.org/10.1016/j.cub.2019.09.056.

Gu, Liuqi, and James R. Walters. 2017. "Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details." *Genome Biology and Evolution* (England) 9 (9): 2461–76. https://doi.org/10.1093/gbe/evx154.

Hof, A.E.V.; Van't Hof, A.E.; Nguyen, P.; Dalíková, M.; Edmonds, N.; Marec, F.; Saccheri, I.J. Linkage map of the peppered moth, Biston betularia (Lepidoptera, Geometridae): A model of industrial melanism. Heredity **2013**, 110, 283–295.

Huylmans, Ann Kathrin, Ariana Macon, and Beatriz Vicoso. 2017. "Global Dosage Compensation Is Ubiquitous in Lepidoptera, but Counteracted by the Masculinization of the Z Chromosome." *Molecular Biology and Evolution* (United States) 34 (10): 2637–49. https://doi.org/10.1093/molbev/msx190.

Kiuchi, Takashi, Hikaru Koga, Munetaka Kawamoto, et al. 2014. "A Single Female-Specific piRNA Is the Primary Determiner of Sex in the Silkworm." *Nature* (England) 509 (7502): 633–36. https://doi.org/10.1038/nature13315.

Lau, Alyssa C, and Györgyi Csankovszki. 2015. "Balancing up and Downregulation of the C. Elegans X Chromosomes." *Genome Architecture and Expression* 31 (April): 50–56. https://doi.org/10.1016/j.gde.2015.04.001.

Lucchesi, John C., William G. Kelly, and Barbara Panning. 2005. "Chromatin Remodeling in Dosage Compensation." *Annual Review of Genetics* (United States) 39: 615–51. https://doi.org/10.1146/annurev.genet.39.073003.094210.

Mora, Pablo, Monika Hospodářská, Anna Chung Voleníková, et al. 2024. "Sex-Biased Gene Content Is Associated with Sex Chromosome Turnover in Danaini Butterflies." *Molecular Ecology* 33 (24): e17256. https://doi.org/10.1111/mec.17256.

Muyle, Aline, Doris Bachtrog, Gabriel A. B. Marais, and James M. A. Turner. 2021. "Epigenetics Drive the Evolution of Sex Chromosomes in Animals and Plants." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1826): 20200124. https://doi.org/10.1098/rstb.2020.0124.

Pokorná, Martina, Marie Altmanová, and Lukáš Kratochvíl. 2014. "Multiple Sex Chromosomes in the Light of Female Meiotic Drive in Amniote Vertebrates."

Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology (Netherlands) 22 (1): 35–44. https://doi.org/10.1007/s10577-014-9403-2.

Sahara, Ken, Atsuo Yoshido, and Walther Traut. 2012a. "Sex Chromosome Evolution in Moths and Butterflies." *Chromosome Research* 20 (1): 83–94. https://doi.org/10.1007/s10577-011-9262-z.

Sahara, Ken, Atsuo Yoshido, and Walther Traut. 2012b. "Sex Chromosome Evolution in Moths and Butterflies." *Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* (Netherlands) 20 (1): 83–94. https://doi.org/10.1007/s10577-011-9262-z.

Samata, Maria, and Asifa Akhtar. 2018. "Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs." *Annual Review of Biochemistry* (United States) 87 (June): 323–50. https://doi.org/10.1146/annurev-biochem-062917-011816.

Traut, W., and F. Marec. 1996. "Sex Chromatin in Lepidoptera." *The Quarterly Review of Biology* (United States) 71 (2): 239–56. https://doi.org/10.1086/419371.

Traut, W., K. Sahara, and F. Marec. 2007. "Sex Chromosomes and Sex Determination in Lepidoptera." *Sexual Development : Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation* (Switzerland) 1 (6): 332–46. https://doi.org/10.1159/000111765.

Traut, Walther, Heiko Vogel, Gernot Glöckner, Enno Hartmann, and David G. Heckel. 2013. "High-Throughput Sequencing of a Single Chromosome: A Moth W Chromosome." *Chromosome Research* 21 (5): 491–505. https://doi.org/10.1007/s10577-013-9376-6.

Veitia, Reiner A. 2005. "Gene Dosage Balance: Deletions, Duplications and Dominance." *Trends in Genetics : TIG* (England) 21 (1): 33–35. https://doi.org/10.1016/j.tig.2004.11.002.

Walters, James R., and Thomas J. Hardcastle. 2011. "Getting a Full Dose? Reconsidering Sex Chromosome Dosage Compensation in the Silkworm, Bombyx Mori." *Genome Biology and Evolution* (England) 3: 491–504. https://doi.org/10.1093/gbe/evr036.

Walters, James R., Thomas J. Hardcastle, and Chris D. Jiggins. 2015. "Sex Chromosome Dosage Compensation in Heliconius Butterflies: Global yet Still Incomplete?" *Genome Biology and Evolution* (England) 7 (9): 2545–59. https://doi.org/10.1093/gbe/evv156.

Wright, Alison E., Rebecca Dean, Fabian Zimmer, and Judith E. Mank. 2016. "How to Make a Sex Chromosome." *Nature Communications* 7 (1): 12087. https://doi.org/10.1038/ncomms12087.

Wright, Charlotte J., Lewis Stevens, Alexander Mackintosh, Mara Lawniczak, and Mark Blaxter. 2024. "Comparative Genomics Reveals the Dynamics of Chromosome Evolution in Lepidoptera." *Nature Ecology & Evolution* 8 (4): 777–90. https://doi.org/10.1038/s41559-024-02329-4.

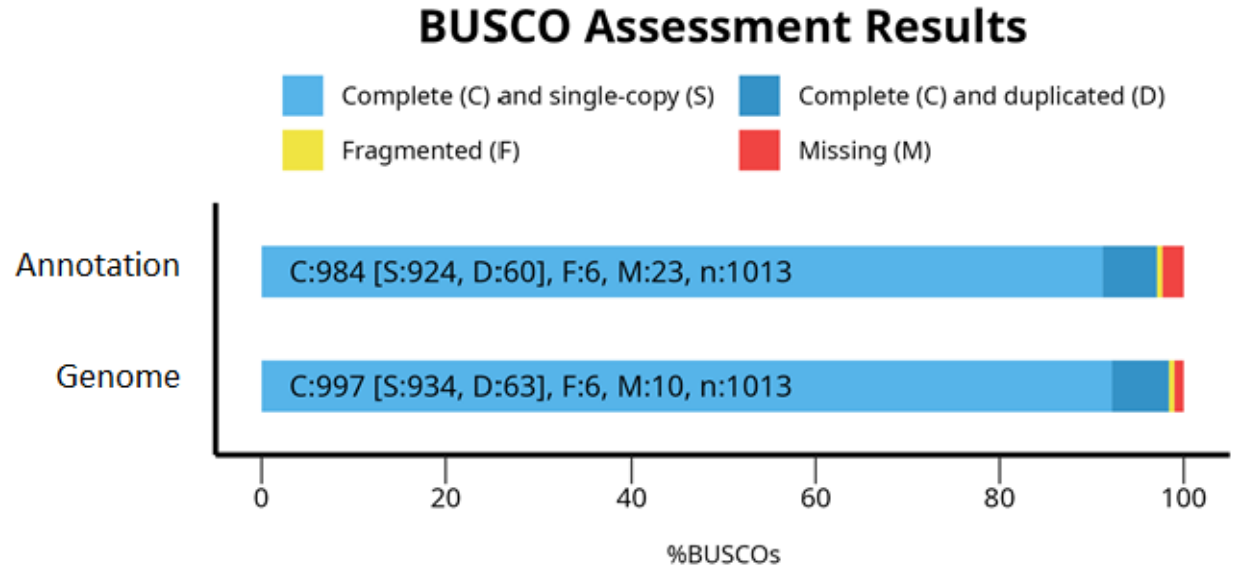
Yoshido, A., F. Marec, and K. Sahara. 2016. "The Fate of W Chromosomes in Hybrids between Wild Silkmoths, Samia Cynthia Ssp.: No Role in Sex Determination and Reproduction." *Heredity* 116 (5): 424–33. https://doi.org/10.1038/hdy.2015.110.

Zimmer, Frederic, Annika Maria Fox, Qiaowei Pan, et al. 2025. "Convergent Evolution of H4K16ac-Mediated Dosage Compensation Shapes Sex-Dependent Lifespan in a ZW Species." *bioRxiv*, January 1, 2025.04.10.648168. https://doi.org/10.1101/2025.04.10.648168.

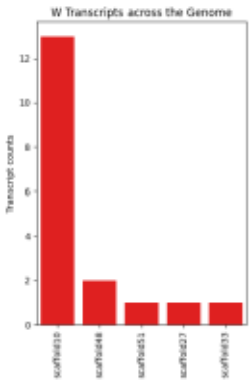
Supplementary Information

Statistic	Contig-level	Scaffold-level
N50 length (bp)	2569677	9121241
N80 length (bp)	1142567	4614841
Longest contig/scaffold length (bp)	9193191	30143683
Number of contigs/ scaffolds	447	220
Total size (bp)	453825874	453869025

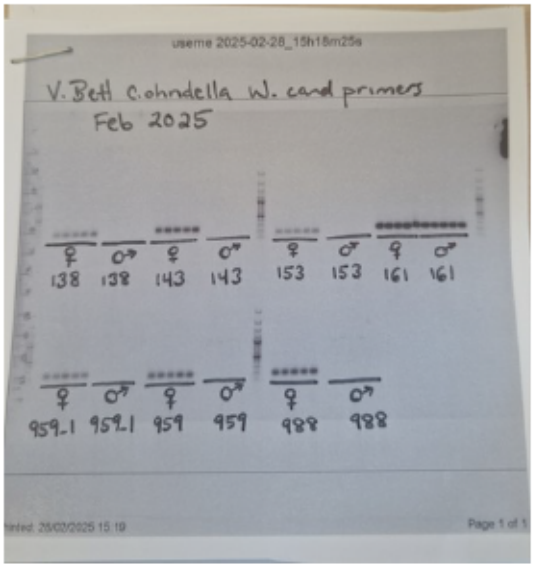
Table S1: Genome assembly statistics of female *Cameraria ohridella*



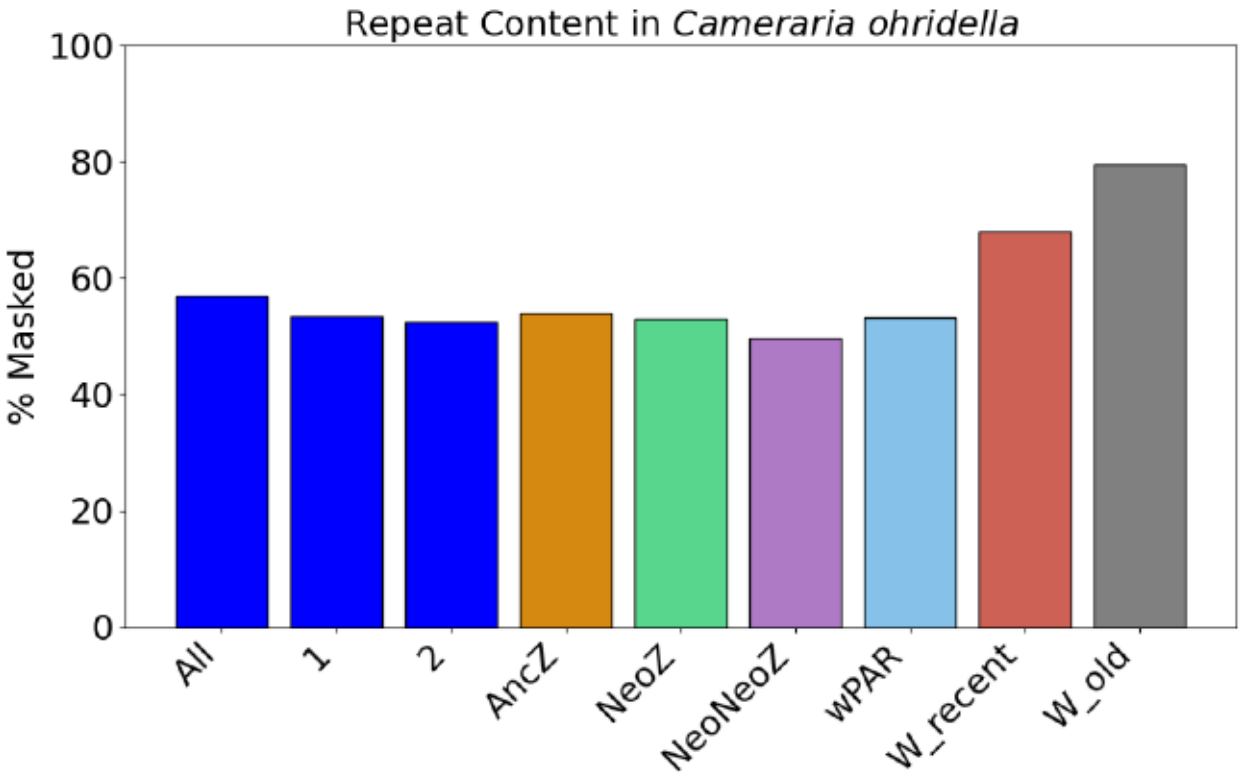
supplementary fig. 1: BUSCO assessment statistic results of genome assembly and annotation of female *C. ohridella*



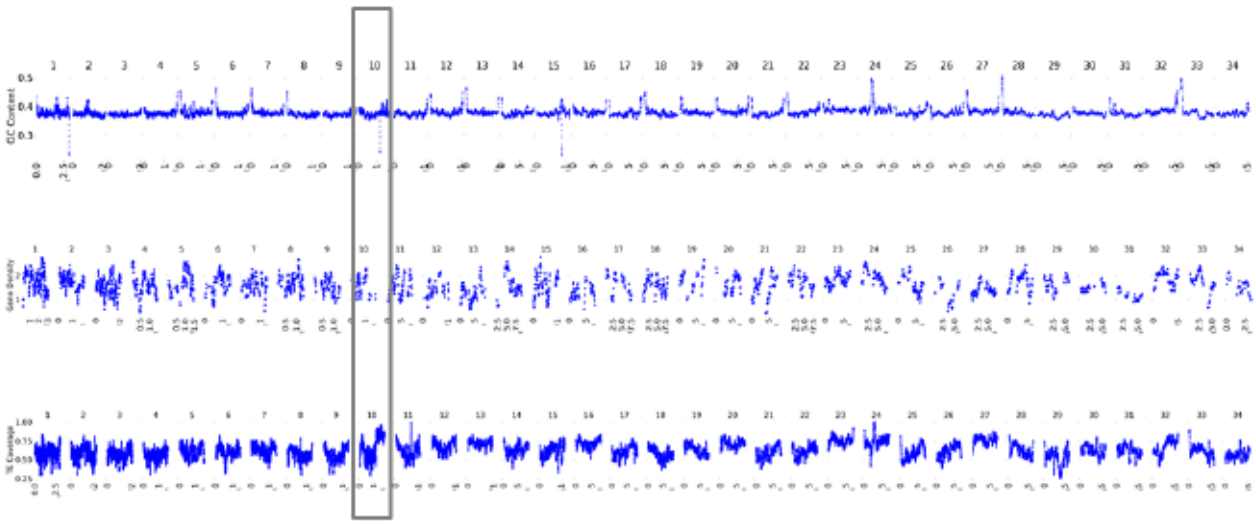
supplementary fig. 2: Distribution of putative W-linked transcripts (identified using K-mer pipeline approach) across the genome.



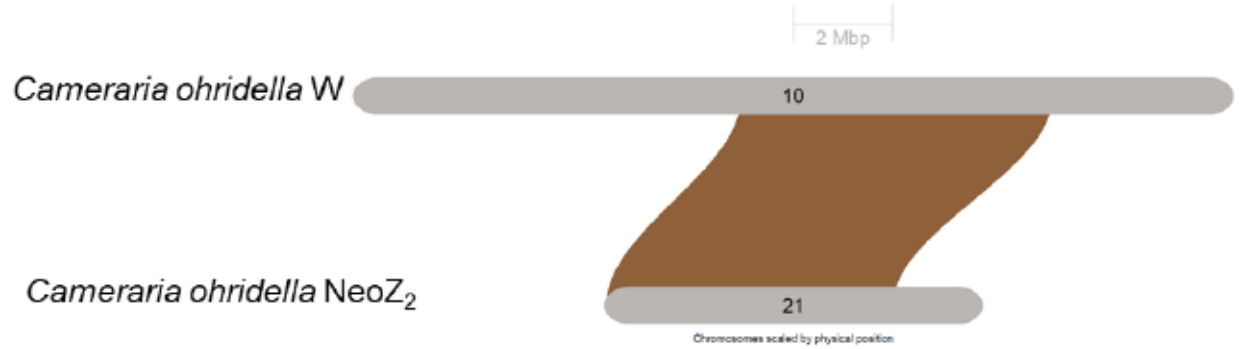
supplementary fig. 3: PCR for selected putative W-linked sequences (seven specific fragment sequences) conducted in five females and five males.



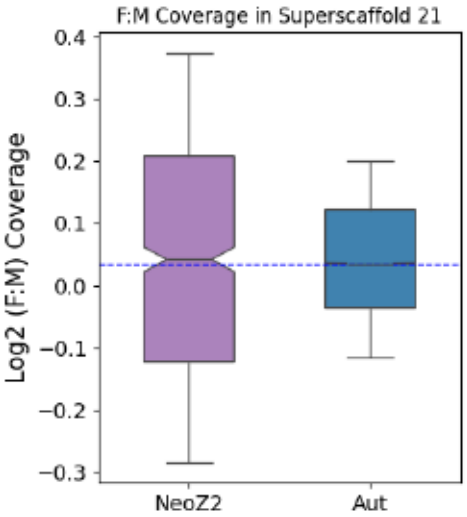
supplementary fig. 4: Repeat masked in across the genome and selected scaffolds and regions of the sex chromosomes in *Cameraria ohridella*



supplementary fig. 5: **GC content, gene density and repeat content distribution across the genome of *Cameraria ohridella***

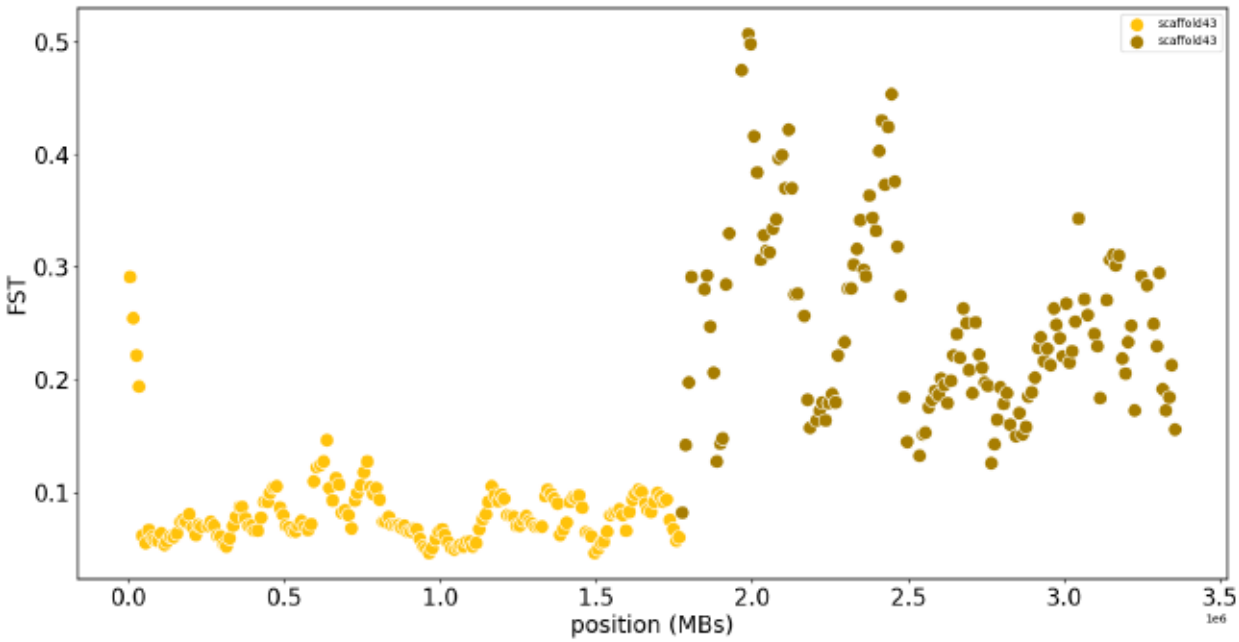


supplementary fig. 6: **Synteny of superscaffold 10 and superscaffold 21**

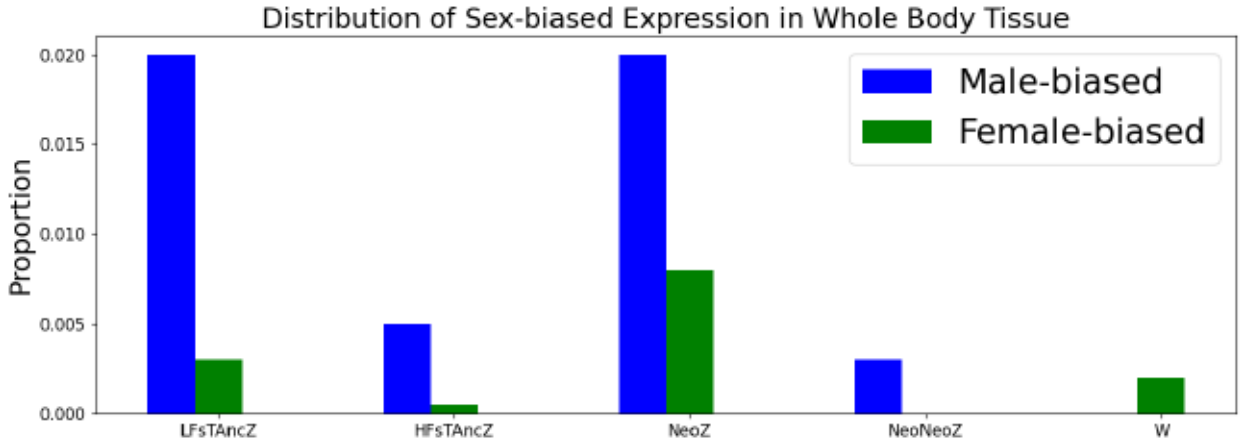


supplementary fig. 7: **Female to male coverage between autosome to neo-Z2 in *Cameraria ohridella***

supplementary fig. 8: Homology of genes in first stratum (S0) in superscaffold 10 identified through an all-vs-all BLAST analysis, filtering for reciprocal best hits with a minimum sequence identity of 50 bp.

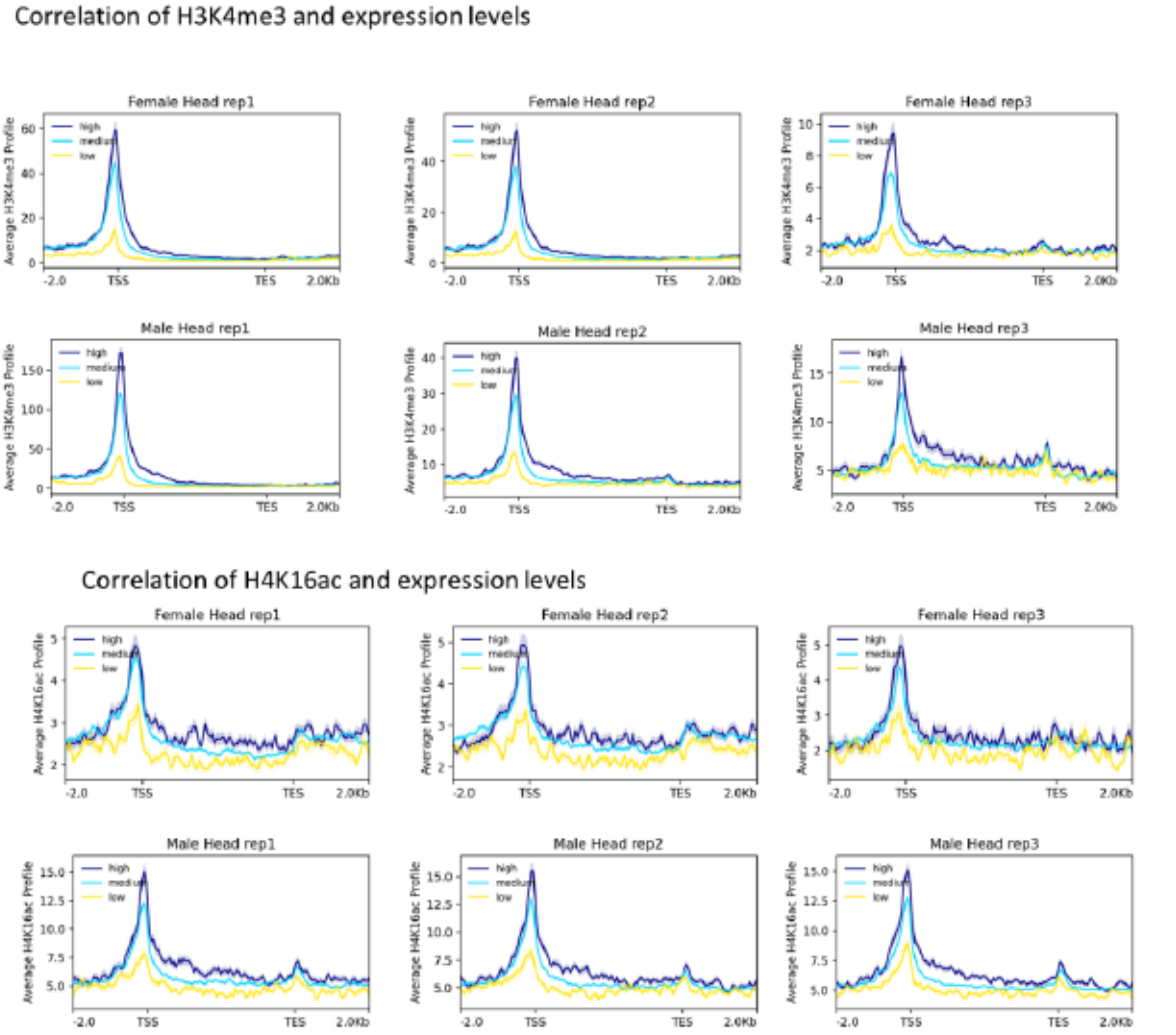


supplementary fig. 9: Fst levels in putative W scaffold (scaffold43) that had some genes homologous to HighFst regions of the ancestralZ

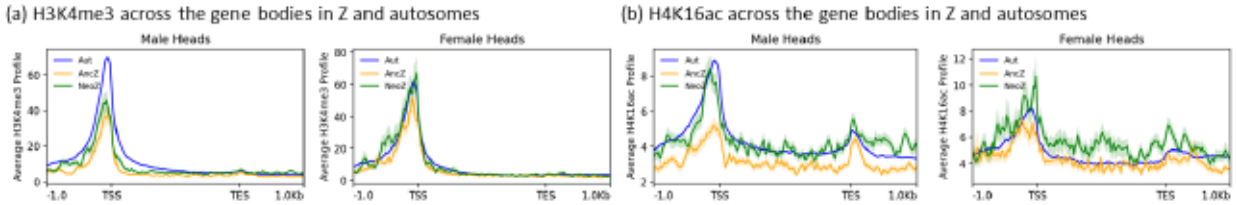


supplementary fig. 10: **Overall sex-biased distribution in the Z chromosome (ancestralZ (LowFST and HighFST regions), neo-Z_1(NeoZ) and neo-Z_2) and also in the W-specific region of superscaffold 10.**

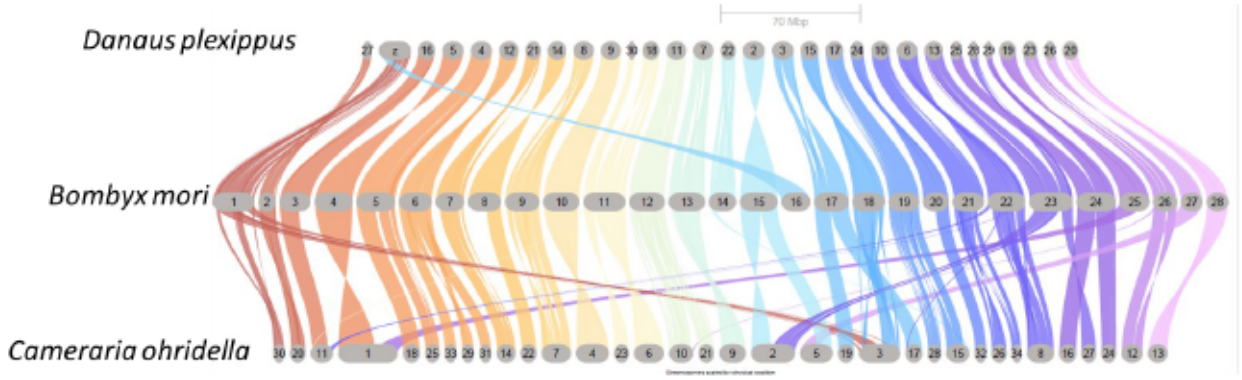
supplementary fig. 11: **Patterns of dosage balance between the Z chromosome and autosome in whole body tissue following exclusion of sex-biased expression.**



supplementary fig. 12: **Correlation of enrichment of histone modifications and expression levels**



supplementary fig. 13: Average profile enrichment of active histone modification H3K4me3 (a) and H4K16ac (b) in the Z chromosome and autosomes of head tissues in males and females



supplementary fig. 14: Synteny among *Danaus plexippus*, *Bombyx mori* and *Cameraria ohridella*

Chapter 5: Thesis discussion

In this thesis, we examined the evolutionary dynamics of sex chromosome differentiation, with a particular focus on the extent of dosage compensation following gene loss in the heterogametic sex and the potential chromatin-based mechanisms underlying this process. We also characterized sex-biased gene expression and investigated its regulation through histone modifications. Furthermore, we showed that the *Artemia franciscana* genome is highly repetitive, with many genes containing transposable elements within their introns. Our analyses revealed that enrichment of histone modifications associated with constitutive heterochromatin correlated with gene expression levels.

Invertebrate epigenetic regulation: Artemia adds diversity to an already complex picture

Epigenetic regulation has emerged as a fundamental piece in the puzzle that is the genotype to phenotype map. Among the many epigenetic processes, histone modifications are among the most extensively studied and well characterized. Modifications defined as "active" or "repressive" in model species are often assumed to be so in distant groups. One of the best characterized modifications is H3K9me3, a marker of heterochromatin typically associated with low transcriptional output. To our (initial) surprise, in our data, both H3K9me3 and H3K9me2 were enriched in highly expressed genes. Our interpretation of this pattern has been greatly shaped by work in Drosophila, where heterochromatic genes require these repressive marks to suppress the expression of transposable elements (TEs) present across their gene bodies (as this spurious transcription interferes with the expression of the genes themselves) (Ninova et al. 2019). As shown in Chapter 2, the genome of *Artemia franciscana* is rich in TEs, and we speculate a similar role of heterochromatic marks in facilitating gene expression through intronic TE repression. This enrichment in introns suggests that these heterochromatic histone marks to suppress spurious expression of transposable elements, thereby maintaining optimal gene transcription level. Since all our RNA-seq libraries were poly-A selected, the resulting data was not ideal for inferring the expression levels of TEs across the genome, which would have allowed for a direct test of this hypothesis. In the future, poly-A-independent RNA-seq data could be used for this purpose, and more generally to infer the complex interplay between histone marks, TEs and gene expression regulation.

A significant but small contribution of chromatin landscape to sexual dimorphism

Differences in gene expression regulation between males and females are thought to underlie much of sexual dimorphism, and are typically maintained over much of the life cycle, which could be consistent with substantial sex-specific chromatin regulation. While sex-biased genes have been identified across diverse taxa, the contribution of chromatin dynamics to sex-biased gene expression has been scarcely explored, and existing studies offer limited evidence for a direct causal role of chromatin in shaping these expression patterns. This gap is striking given that chromatin states are well recognized as key regulators of developmental gene expression programs in animals (Lindeman et al. 2010). In Chapters 3, we examined sex-biased gene expression patterns in the heads and gonads of *Artemia franciscana* and investigated their association with histone modifications. Our analysis of epigenetic regulation of sex-biased gene expression to histone modifications, and found an association between sex-biased gene expression and contrasting chromatin states in males and females. However, while statistically significant, this association was relatively weak, with many sex-biased genes showing concordant chromatin states in males and females. Why would chromatin landscape contribute relatively little, when it is known to affect other instances of genome plasticity? One possibility

is that other, unsampled, histone marks contribute to sex-specific regulation, or that our classification of open and closed chromatin was not precise enough to fully uncover its functional role. Using broad-scale measures of chromatin openness, such as those obtained from ATAC-seq data, could in the future shed light on these possibilities. However, by sampling adults, we also only obtained a late snapshot of how sexual dimorphism is set up and maintained. Studying sex-specific chromatin landscape throughout early development will be needed before its full contribution to sexual dimorphism can be inferred.

A more extensive developmental sampling of sex-specific chromatin states may also in the future shed light on one of the questions that our lab has been trying to solve: how is sex determined in Artemia? Whereas DNA methylation regulates key sex-determining genes like *SRY* in mammals during gonadal development (Nishino et al. 2004), the molecular cascade governing sex determination and sexual differentiation in *Artemia franciscana* remains unresolved. Our lab is currently undertaking a broad search to identify candidate genes potentially involved in this process. An intriguing question is whether female-specific methylation patterns on the W chromosome play a regulatory role, and how such epigenetic modifications might influence sex determination and differentiation in this species.

ZW systems can have global compensation, but the mechanisms vary

The evolution of heterogametic sex chromosomes causes gene dose differences between sexes: in ZW systems, males retain two copies of Z-linked genes while females carry only one. This imbalance can disrupt interactions between Z-linked and autosomal loci in females, so dosage compensation evolves to restore ancestral expression levels and maintain proper protein stoichiometry. Early findings from ZW systems, including birds, snakes and certain fishes, supported the view that these systems lack global compensation or limit it to only a set of dosage-sensitive genes. Why many ZW systems lack global dosage compensation remains unclear. Gu and Walters (2017) proposed several, potentially interacting, mechanisms underlying differences in dosage compensation between XY and ZW systems (Gu and Walters 2017). The first hypothesis attributes this pattern to male mutation bias since oogenesis entails fewer DNA replications than spermatogenesis, degeneration of the W chromosome proceeds more slowly than that of the Y, favoring localized rather than chromosome-wide compensation on the Z chromosome. Additionally, the reduced effective population size of the Z chromosome, resulting from greater variance in male reproductive success, may hinder adaptation relative to the X in XY systems (Gu and Walters 2017). Finally, male-biased Z transmission could favor fixation of male-beneficial, sexually antagonistic genes, reinforcing sex-biased expression instead of balanced compensation (Huylmans et al. 2017).

In Chapters 2 and 4, we investigated patterns of dosage compensation within the differentiated regions of the Z chromosome in *Artemia franciscana* and in *Cameraria ohridella*, the latter of which possesses three distinct Z chromosome regions defined by their evolutionary trajectories. In *A. franciscana*, our analyses revealed that gene expression levels in the differentiated Z region are comparable to those observed on the autosomes across both male and female heads and gonads, thereby indicating the presence of complete dosage compensation. By contrast, in *C. ohridella*, we found that dosage compensation varies across the distinct regions of the Z chromosome: an ancient Z arm shared with all Lepidoptera, and two additional neo-Z chromosomes of different ages (neo-Z_1 and neo-Z_2). Differences were also evident between head and whole-body tissues. In heads, dosage balance (Z = autosomes in both sexes) was complete on the (older) neo-Z_1 but incomplete (Z<autosomes only in females) on the neo-Z_2. The ancestral Z (anc-Z) showed overall balance between the sexes but incomplete dosage compensation (Z<autosomes in both sexes), but female-biased expression in regions that recently duplicated to the W chromosome. In whole-body tissues, all Z chromosome regions

shifted to lower female:male ratios of expression, likely due to the inclusion of gonads, which often show unusual patterns of expression of sex-linked genes. These results highlight the complexity of dosage compensation evolution, but also the potential contribution of neo-sex chromosomes to our understanding of its progression. For instance, neo-Z_2 appears to be at an intermediate stage, with partial balancing of expression (not fully balancing male and female transcriptional levels), while neo-Z_1 seems to have reached full expression balance.

Neo-sex chromosomes can shed important light on the evolution of dosage compensation

The evolution of dosage compensation on neo-sex chromosomes followed strikingly diverse regulatory trajectories across lineages such as Drosophila, Lepidoptera and Coleoptera (Kalita and Keller Valsecchi 2025). In *Drosophila miranda*, the second neoX, which was acquired ~1 Mya, already shows partial dosage compensation, with intermediate *Msl-2* binding and H4K16ac enrichment, reflecting its gradual acquisition of *MSL* sites via transposable elements (Ellison and Bachtrog 2013). The older neo X chromosome, which has existed for over 10 My, is already fully compensated. Similarly, *Tribolium confusum* has a fully compensated neoX (~20 Mya) , equivalent to the ancestral X (Bracewell et al. 2024). This suggests that, in Diptera and Coleoptera, neo X chromosomes consistently co-opt the pre-existing mechanism of compensation, and that the process is complete in a few million years. In Lepidoptera, the regulation of neo-Z chromosomes is often at odds with that of the ancestral Z. The neo-Z chromosome of *Cydia pomonella* is an exception, as it shows partial downregulation, mirroring the reduced expression of the ancestral Z (Gu et al. 2017). Among Danaini butterflies (*Idea leuconoe*, *Lycorea halia*, *Danaus plexippus*), the ancestral Z is sex-balanced but has lower expression level relative to autosomes, while the neoZ is completely compensated (Kalita and Keller Valsecchi 2025). Notably, the neoZ of *I. leuconoe* and *D. plexippus* shares a common origin (~26 Mya), whereas that of *L. halia* arose independently via fusion to a different autosomal chromosome (Kalita and Keller Valsecchi 2025; Bracewell et al. 2024).
In *C. ohridella*, H4K16ac displayed a striking dichotomy between the two segments of the Z chromosome: in females, it was enriched on the $NeoZ_1$, while in males the ancestral Z region was depleted of this mark. Similarly, a dichotomous pattern of H4K16ac has been reported in the Monarch butterfly, despite differences in the chromosomal origins of the neo-Z (fusion of chromosome 16 with the ancestral Z in Monarch, chromosome 18 with the ancestral Z in *C. ohridella* (based on *Bombyx mori* chromosome nomenclature). These results show that differential chromatin regulation of neo- and ancestral-Z arms, and consequent transcriptional differences, are widespread in Lepidoptera. Why might this be? One difference between Diptera and Lepidoptera is the mechanism of compensation of the ancestral X/Z: upregulation of the male X in Diptera versus downregulation of the male Z in Lepidoptera. As neo-X chromosomes acquire binding sites for the ancient mechanisms of dosage compensation, they become immediately balanced with the autosomal expression, which should be favored by selection. On the other hand, a neo-Z targeted by the ancient mechanism of dosage compensation will become down-regulated and out of balance with autosomes. Neo-Z chromosomes may consequently require novel mechanisms to balance their expression (Toups and Vicoso 2025).

Striking molecular convergence between Artemia and Drosophila

Dosage compensation relies on epigenetic modulation of chromatin accessibility. In *Drosophila*, this occurs through transcriptional upregulation of the X chromosome in males; in mammals, through X-chromosome inactivation in females; and in *C. elegans* hermaphrodites, through partial downregulation of both X chromosomes. In mammals, inactivation is mediated

by the long noncoding RNAs (lncRNA) *Xist*, which acts in cis to recruit protein complexes that will promote accumulation of repressive histone modifications H3K27me3, H2AK119ub, and DNA methylation leading to chromatin compaction and transcriptional silencing and this occurs early during female development (Galupa and Heard 2018). In *Drosophila*, *roX* RNA associates with the male-specific lethal (MSL) complex, which acetylates H4K16 to enhance transcription from the male X chromosome. X-chromosome regulation begins also early in development, with male-specific Msl-2 enabling full assembly of the MSL complex (Conrad and Akhtar 2012). The MSL complex targets the X chromosome in two steps: it first binds high affinity sites (*HAS*) which are *GA*-rich, then spreads to actively transcribed genes through *Msl-3*'s affinity for H3K36me3 (Toups and Vicoso 2025).

In *Artemia franciscana*, we found significant enrichment of H4K16ac on Z-specific regions of the Z chromosome in females, across both head and gonad tissues consistent with dosage compensation via upregulation of the female Z. We also identified a candidate AT-rich motif enriched specifically in the differentiated Z region, but absent from pseudoautosomal and autosomal regions. This motif often co-occurs with H4K16ac peaks, suggesting a role in recruiting dosage compensation machinery. However, because some Z-linked genes lack the motif and show no expression differences relative to motif-containing genes, spreading of compensation likely occurs through an unknown mechanism, enabling broad upregulation across the Z. H4K16ac-mediated dosage compensation has also been reported in green anole lizards, where males show X-chromosome enrichment directed by the lncRNA *MAYEX*, which associates with histone H4 and a putative acetyltransferase (G1KD42) (Tenorio et al. 2024). This enzyme differs from the histone acetyltransferase *MOF* (Male absent on the first) in fruit flies, which interacts with *roX* RNA to promote chromosome-wide dosage compensation. These findings suggest despite mechanistic diversity, dosage compensation converges on H4K16ac deposition as a common strategy for balancing sex chromosome expression. Long noncoding RNAs which are the key regulators of dosage compensation; *XIST* in humans, *RSX* in opossums, and *MAYEX* in anoles display striking sex-specific expression patterns in somatic tissues (Tenorio et al. 2024). In *A. franciscana*, we did not detect sex-specific expression in the Z-linked region (200 bp windows, in adult head tissues), however, our analysis relied on mapping using non-specific stranded RNA-seq data and was limited to a single adult developmental stage. A recent study found that *Drosophila* MSL complex subunits are neither sex-biased nor activated during embryonic dosage compensation in *Artemia* (Zimmer et al. 2025), implying that dosage compensation may be governed by alternative, as-yet-unidentified mechanisms.

References

Bracewell, Ryan, Anita Tran, Kamalakar Chatla, and Doris Bachtrog. 2024. "Sex and Neo-Sex Chromosome Evolution in Beetles." *PLOS Genetics* 20 (11): e1011477. https://doi.org/10.1371/journal.pgen.1011477.

Brown, Emily J., and Doris Bachtrog. 2014. "The Chromatin Landscape of Drosophila: Comparisons between Species, Sexes, and Chromosomes." *Genome Research* (United States) 24 (7): 1125–37. https://doi.org/10.1101/gr.172155.114.

Brown, Emily J., Alison H. Nguyen, and Doris Bachtrog. 2020. "The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide." *Molecular Biology and Evolution* (United States) 37 (10): 2808–24. https://doi.org/10.1093/molbev/msaa082.

Dixon, Groves, and Mikhail Matz. 2022. "Changes in Gene Body Methylation Do Not Correlate with Changes in Gene Expression in Anthozoa or Hexapoda." *BMC Genomics* (England) 23 (1): 234. https://doi.org/10.1186/s12864-022-08474-z.

Ellison, Christopher E., and Doris Bachtrog. 2013. "Dosage Compensation via Transposable

Element Mediated Rewiring of a Regulatory Network." *Science (New York, N.Y.)* (United States) 342 (6160): 846–50. https://doi.org/10.1126/science.1239552.
Gu, Liuqi, and James R. Walters. 2017. "Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details." *Genome Biology and Evolution* (England) 9 (9): 2461–76. https://doi.org/10.1093/gbe/evx154.
Gu, Liuqi, James R. Walters, and Douglas C. Knipple. 2017a. "Conserved Patterns of Sex Chromosome Dosage Compensation in the Lepidoptera (WZ/ZZ): Insights from a Moth Neo-Z Chromosome." *Genome Biology and Evolution* 9 (3): 802–16. https://doi.org/10.1093/gbe/evx039.
Gu, Liuqi, James R. Walters, and Douglas C. Knipple. 2017b. "Conserved Patterns of Sex Chromosome Dosage Compensation in the Lepidoptera (WZ/ZZ): Insights from a Moth Neo-Z Chromosome." *Genome Biology and Evolution* (England) 9 (3): 802–16. https://doi.org/10.1093/gbe/evx039.
Huylmans, Ann Kathrin, Ariana Macon, and Beatriz Vicoso. 2017. "Global Dosage Compensation Is Ubiquitous in Lepidoptera, but Counteracted by the Masculinization of the Z Chromosome." *Molecular Biology and Evolution* (United States) 34 (10): 2637–49. https://doi.org/10.1093/molbev/msx190.
Kalita, Agata Izabela, and Claudia Isabelle Keller Valsecchi. 2025. "Dosage Compensation in Non-Model Insects - Progress and Perspectives." *Trends in Genetics : TIG* (England) 41 (1): 76–98. https://doi.org/10.1016/j.tig.2024.08.010.
Marshall, H., M. T. Nicholas, J. S. van Zweden, et al. 2023. "DNA Methylation Is Associated with Codon Degeneracy in a Species of Bumblebee." *Heredity* (England) 130 (4): 188–95. https://doi.org/10.1038/s41437-023-00591-z.
Mendoza, Alex de, Ryan Lister, and Ozren Bogdanovic. 2020. "Evolution of DNA Methylome Diversity in Eukaryotes." *Reading DNA Modifications* 432 (6): 1687–705. https://doi.org/10.1016/j.jmb.2019.11.003.
Nishino, Koichiro, Naoko Hattori, Satoshi Tanaka, and Kunio Shiota. 2004. "DNA Methylation-Mediated Control of Sry Gene Expression in Mouse Gonadal Development." *The Journal of Biological Chemistry* (United States) 279 (21): 22306–13. https://doi.org/10.1074/jbc.M309513200.
Okwaro, Louis A., and Judith Korb. 2023. "Epigenetic Regulation and Division of Labor in Social Insects." *Current Opinion in Insect Science* (Netherlands) 58 (August): 101051. https://doi.org/10.1016/j.cois.2023.101051.
Qureshi, Irfan A., and Mark F. Mehler. 2010. "Genetic and Epigenetic Underpinnings of Sex Differences in the Brain and in Neurological and Psychiatric Disease Susceptibility." *Progress in Brain Research* (Netherlands) 186: 77–95. https://doi.org/10.1016/B978-0-444-53630-3.00006-3.
Tenorio, Mariela, Samantha Cruz-Ruiz, Sergio Encarnación-Guevara, et al. 2024. "MAYEX Is an Old Long Noncoding RNA Recruited for X Chromosome Dosage Compensation in a Reptile." *Science* 385 (6715): 1347–54. https://doi.org/10.1126/science.adp1932.
Toups, Melissa A., and Beatriz Vicoso. 2025. "Insect Sex Chromosome Evolution: Conservation, Turnover, and Mechanisms of Dosage Compensation." *Current Opinion in Insect Science*, July 12, 101411. https://doi.org/10.1016/j.cois.2025.101411.
Walters, James R., and Thomas J. Hardcastle. 2011. "Getting a Full Dose? Reconsidering Sex Chromosome Dosage Compensation in the Silkworm, Bombyx Mori." *Genome Biology and Evolution* (England) 3: 491–504. https://doi.org/10.1093/gbe/evr036.
Yen, Angela, and Manolis Kellis. 2015. "Systematic Chromatin State Comparison of Epigenomes Associated with Diverse Properties Including Sex and Tissue Type." *Nature Communications* (England) 6 (August): 7973. https://doi.org/10.1038/ncomms8973.

Zimmer, Frederic, Annika Maria Fox, Qiaowei Pan, et al. 2025. "Convergent Evolution of H4K16ac-Mediated Dosage Compensation Shapes Sex-Dependent Lifespan in a ZW Species." *bioRxiv*, January 1, 2025.04.10.648168. https://doi.org/10.1101/2025.04.10.648168.