Evolution of sex chromosomes, sex determination and asexuality in *Artemia* brine shrimp

by
Marwan Elkrewi
February, 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

Thesis defense Committee:
Maximilian Jösch, Chair
Beatriz Vicoso, Supervisor
Lora B. Sweeney, Internal committee member
Christoph Haag, External committee member

The thesis of Marwan Elkrewi, titled “Evolution of sex chromosomes, sex determination and asexuality in *Artemia* brine shrimp”, is approved by:

Supervisor: Beatriz Vicoso, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Lora B. Sweeney, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Christoph Haag, CEFE - CNRS Montpellier, France

Signature: _____

Defense Chair: Maximilian Jösch, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

ISBN: 978-3-99078-053-4

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: _____

Signed page is on file

Abstract

Crustaceans are a large group of arthropods with a great diversity of species and different types of sex determination systems and reproductive modes (Subramoniam, 2017). This makes them a great model for exploring the evolution of sex chromosomes and sexual dimorphism and investigating the evolutionary mechanisms driving and maintaining the diversity of reproductive systems. Within this taxon, Brine shrimp of the genus *Artemia*, a branchiopod crustacean, are well suited for such explorations, as they have both highly dimorphic traits and closely related sexual and asexual species. Although brine shrimp are known to have ZW sex chromosomes (Bowen, 1963; Parraguez et al., 2009), the sex chromosomes are still not well characterized at the genomic level, the sex-determination gene is unknown, and it is still unclear whether the same sex chromosomes as shared by the different species.

The first part of this thesis was to characterize the Z and W chromosomes in *Artemia* using an array of methods, from generating multiple chromosome and contig level genome assemblies to identifying W-linked scaffolds and transcripts in multiple species using k-mer based approaches.

The second part tackles the conservation of the cell type specific regulatory pathways in the female reproductive system between *Artemia* and *Drosophila*, and the expression of the Z-specific region throughout meiosis using single-nucleus RNA-seq data. Our results show that germline cells lack dosage compensation, with a subset of cells showing evidence of extreme repression of the Z chromosome.

With multiple sexual species and several asexual lineages of parthenogenetic females that produce rare males at low frequencies, Brine shrimp present the perfect opportunity to explore the transition to asexuality and shed light on the prerequisites and repercussions of the form of modified meiosis maintaining the asexual lineages. The last chapter is an investigation of the molecular pathways involved in asexual reproduction in *Artemia* using newly generated single nucleus RNAseq and WGS data and previously published data.

Acknowledgments

I would like to thank my PhD supervisor, Beatriz Vicoso, for the amazing experience, all the valuable knowledge, continued support and guidance. Looking back, I would not change anything about my time in her lab. I would also like to thank my committee members Lora B. Sweeney and Christoph Haag for the valuable input at the different stages of my PhD, and the ISTA Scientific computing team and the Vienna Biocenter Sequencing facility for technical support.
I am extremely grateful to Daria Siekhaus for getting me here, believing that I could do this, and for being there throughout the whole process. Many thanks to all the members of the Vicoso group, past and present ones, for sharing the journey with me and making it more joyful, on both the scientific and human levels. For my friends and family, this would not be possible without you, and from my side, no words would ever do you justice.

My PhD work was funded by the Austrian science fund (FWF), as part of the SFB Meiosis consortium (https://sfbmeiosis.org/, grant ID FWF SFB F88-10).

About the Author

Marwan Elkrewi completed his Bachelor of Science in Electrical engineering, with a minor in Philosophy, at the American University of Sharjah, UAE in 2016. He joined Weatherford International in their Abu Dhabi office in the same year as a junior field engineer and then transitioned into the role of Service Quality Engineer. During his time there, he was engaged in some genomics research at NYU Abu Dhabi, while also doing the Stanford Genetics and Genomics certificate. The side endeavors allowed him to secure a PhD position at ISTA, which he started in 2019. During his time as a PhD student, he worked on several aspects of sex-chromosome evolution in various organisms, including schistosomes, *Artemia* brine shrimp, and scorpion flies. His work has been presented in various conferences, including the European Meeting for PhD Students in Evolutionary Biology (EMPSEB28) in 2023, and the Congress of the European Society for Evolutionary Biology (ESEB) in 2022. Some of his work has been published in various journals, including MBE, GBE, genetics, and PLOS genetics.

List of collaborators and publications

Publications:

Elkrewi M, Vicoso B. Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome. PLOS genetics. 2024 Aug 30;20(8):e1011376.

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

Lasne C*, Elkrewi M*, Toups MA, Layana L, Macon A, Vicoso B. The scorpionfly (*Panorpa cognata*) genome highlights conserved and derived features of the peculiar dipteran X chromosome. Molecular Biology and Evolution. 2023 Dec;40(12):msad245.

Kelemen RK, Elkrewi M, Lindholm AK, Vicoso B. Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. Proceedings of the Royal Society B. 2022 Feb 9;289(1968):20211985.

Elkrewi M*, Khauratovich U*, Toups MA*, Bett VK, Mrnjavac A, Macon A, Fraisse C, Sax L, Huylmans AK, Hontoria F, Vicoso B. ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp. Genetics. 2022 Oct 1;222(2):iyac123.

Elkrewi M*, Moldovan MA*, Picard MA, Vicoso B. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. Molecular Biology and Evolution. 2021 Dec 138(12):5345-58.

Table of Contents

Abstract v
Acknowledgments vi
About the Author vii
List of collaborators and publications viii
Table of Contents ix
Chapter I: General Introduction 2
Sex chromosome evolution 2
Dosage compensation and the lack of it 3
Meiotic sex chromosome Inactivation 3
Meiosis and its modifications 5
Artemia brine shrimp 7
References 9
Chapter 2: General Abstract 13
Chapter 2.1: ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp 14
Introduction 14
Results 16
Discussion 23
Materials and methods 25
References 31
Supplementary Material 37
Supplementary Datasets 37
Supplementary Figures 38
Supplementary Tables 55
Chapter 2.2: Chromosome-level assembly of *Artemia franciscana* sheds light on sex-chromosome differentiation 58
Abstract 58
Significance 58
Introduction 58
Results and Discussion 59
Materials and Methods 66
References 69
Supplementary Material 72
Supplementary Datasets 72
Supplementary Figures 73
Supplementary Tables 77

Chapter 3: A novel regulatory mechanism for the conserved sex determination gene Doublesex in Artemia brine shrimp 81
Abstract 81
Introduction 81
Results 83
Discussion 91
Materials and Methods 93
References 95
Supplementary material 99
Chapter 4: Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome 104
Abstract 104
Author summary 104
Introduction 105
Results 107
Discussion 115
Methods 118
References 121
Supplementary material 126
Chapter 5: A candidate Z-linked gene for Asexual Reproduction in *Artemia* Brine Shrimp 145
Supplementary material 157
Supplementary tables 157
Supplementary Figures 160
Chapter 6: Thesis discussion 164
Artemia, a new model for studying sex chromosome evolution 164
Some sex determining genes are easier to find than others 165
MSCI: how do we settle a case like *Drosophila* (or *Artemia*)? 166
Why aren't we all asexual? 166
References 168
Appendix 170

CHAPTER 1

General Introduction

Chapter I: General Introduction

This thesis addresses several aspects of the reproductive biology of *Artemia* brine shrimp, from the genes and chromosomes involved in sex determination, to comparisons with lineages that have lost sexual reproduction entirely. Therefore, I start by broadly introducing sex chromosomes and their unusual regulation in somatic and germ tissues, then summarize our knowledge of what modifications are known to be involved in transitions to asexual reproduction. Finally, I briefly describe our model system, which comprises various sexual and asexual lineages of the brine shrimp (Crustacea, genus *Artemia*).

Sex chromosome evolution

Across the tree of life, in organisms with two or more separate sexes, various systems of sex determination have evolved. Those systems fall under three major categories: environmental sex determination (ESD), where the individuals rely on environmental cues to determine sex, cytoplasmic sex determination, where sex is determined by cytoplasmic elements, such as intracellular parasites or mitochondria, and genotypic sex determination (GSD), where the sex is determined by sex specific loci, such as specialized sex chromosomes (1,2). Although many examples of environmental and cytoplasmic sex determination exist in nature, in most organisms, sex is determined genotypically, and sex chromosomes seem to have evolved independently multiple times (2). In the classical scenario, sex chromosomes evolve from a pair of autosomes which acquire a sex determining gene, either by a duplication event or by allele differentiation (Figure 1) (3). In many cases, genes involved in the gonadal differentiation pathways and their interactors are common candidates for acquiring the novel function of initiating ovarian or testicular differentiation (2). After the acquisition of the sex determining loci, the chromosome undergoes a cascade of events. Those events include the loss of recombination in the region with the sex determining gene, potentially due to selection favoring the linkage between the sex determination region and sexually antagonistic alleles, and the accumulation of repeats and deleterious mutations, resulting in the progressive degeneration of the sex-specific chromosome (4). In systems with a male-specific sex chromosome, or male heterogamety, the sex chromosomes are called XY, and in female heterogametic systems, they are called ZW (5). Despite the great progress we have made in understanding the evolution of sex chromosomes in the years since the introduction of the classical model, many questions remain unanswered and many assumptions still lack empirical evidence, especially surrounding the level of conservation of the gonadal differentiation pathways, including the master sex determination gene, and the rates of differentiation and degeneration of sex chromosomes between different lineages (6). While many XY pairs of various ages have been characterized recently, ZW systems remain generally understudied, and a full characterization of the Z and W chromosomes in multiple species is needed to shed light on the parallels and differences between the evolution of the two types of sex chromosomes.

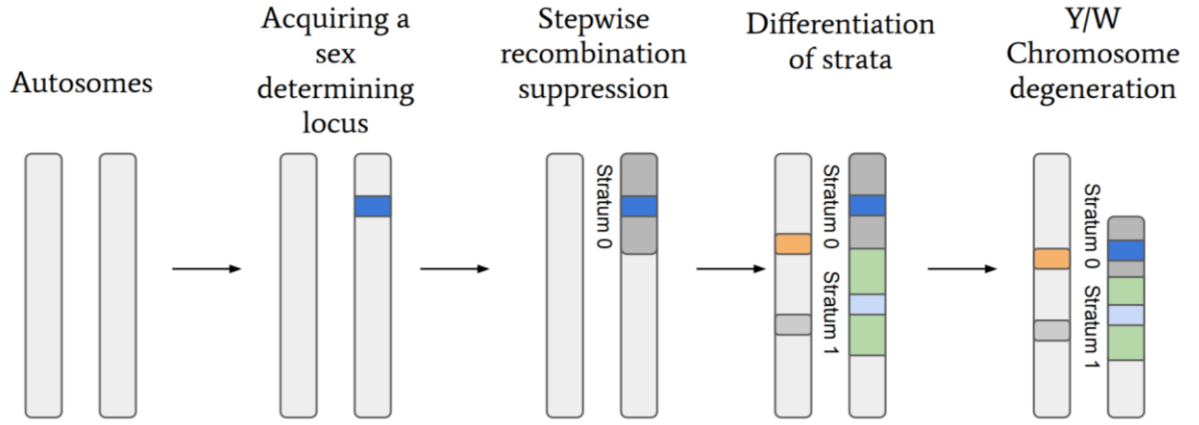


Figure 1: A schematic of the canonical model of sex chromosome evolution.

Dosage compensation and the lack of it

One of the major consequences of having differentiated sex chromosomes is the need to evolve mechanisms to avoid the deleterious effects of having X/Z to autosomes dosage imbalances in the heterogametic sex, driven by the difference in copy numbers (7–9) . Three different solutions to the dosage problem have been observed in nature: doubling the dose of the single copy of the X chromosome (in *Drosophila* males), halving the dose of the two copies of the X chromosome (in *C. elegans* hermaphrodites), and silencing one of the two copies (in female mammals) (10). Another solution involves selectively equalizing the dose of the single-copied essential genes between males and females, which is the case in birds (11,12). However, recent work in Chicken and Platypus suggests that the incomplete dosage compensation observed at the transcriptional-level might be balanced post-transcriptionally (13). It should also be noted that for some groups with ZW chromosomes, such as snakes and some lizards, there is very little evidence of compensation (although again this has only been studied at the transcriptomic level) (14). Why some Z chromosomes do not seem to require a general mechanism of expression balancing is currently unclear.

This diversity in strategies of compensation is mirrored in a diversity of underlying mechanisms, with different genes and pathways having been recruited for this purpose even in flies and mosquitoes (15). However, there are some interesting examples of molecular convergence, suggesting that a limited set of tools can evolve chromosome-wide regulatory functions. The most striking is the specific hyperacetylation of lysine 16 of the histone H4 (H4K16ac) on the X chromosomes of *Drosophila* and the green anole lizards (16). This histone mark is associated with active transcription, and its enrichment on the single X of males of both species is thought to directly cause the doubling of transcription responsible for dosage compensation (16,17). Another convergent pattern is the mediation of dosage compensation by long non-coding RNAs: the long non-coding RNA XIST coats the X chromosome in female mammals and triggers its silencing through a cascade of events that involve acquiring repressive histone marks (18), and the recently discovered MAYEX in reptiles similarly coats the X chromosome (19). However, that takes place in males and results in the upregulation of the single-X by associating with H4K16ac (19).

In many cases, dosage compensation mechanisms are reversed or absent in the germline cells of many species (20,21). As dosage compensation typically involves modifications to the chromatin landscape, its reversal or absence is often associated with the extensive reprogramming during germline development (22). In mammalian female germ cells, the silenced X chromosome is reactivated at the same time as epigenetic reprogramming is taking place (21). In *Drosophila*, evidence suggests that dosage compensation is absent in the primordial germ cells of *Drosophila melanogaster* male embryos (20), and that it shuts down during meiosis in the adult male flies of *Drosophila miranda* (23).

Meiotic sex chromosome Inactivation

During mammalian male meiosis, the failure of the non-recombining regions of the X and Y chromosomes to synapse leads to the accumulation of repressive histone modifications (24,25). Various proteins, including the histone methyltransferase SETDB1 which deposits H3K9me3, mediate the heterochromatization of the sex chromosomes and the formation of a so-called XY body (Figure 2, (26)). The condensed structure presents an obstacle to the transcription machinery, leading to meiotic sex chromosome inactivation (MSCI) (24,27). This process is common to all mammals (28). In terms of its evolutionary significance, it has been postulated that MSCI evolved as a special case from a more general mechanism to protect unpaired DNA from ectopic recombination and insertions (29–31). Another hypothesis is that it evolved to suppress sex-linked meiotic drivers and sex-ratio distorters (32).

Besides mammals, an analogous form of MSCI is also present in Caenorhabditis, with evidence from four different species suggesting the conservation of the mechanism across the genus but not the repressive marks mediating it (33). The evidence from other species is less conclusive. For instance, in *Drosophila*, despite recent evidence suggesting the observed

downregulation of expression of X-linked genes in the germline is caused by a lack of dosage compensation in the male germ cells rather than inactivation (34,23), some recent studies argue for the conservation of MSCI across the *Drosophila* genus (35). There are also conflicting reports in chicken (ZW), where cytogenetic evidence has been used to argue both for and against the presence of MSCI (30,36). In Lepidoptera (ZW), the evidence from two different species suggests the Z is transcriptionally active during meiosis (37). Once again, it is unclear why MSCI arises in some lineages in response to sex chromosome differentiation while meiosis in other lineages seems impervious to their presence. More studies of MSCI in XY and ZW systems are needed before any useful parallels and/or distinctions can be drawn between them.

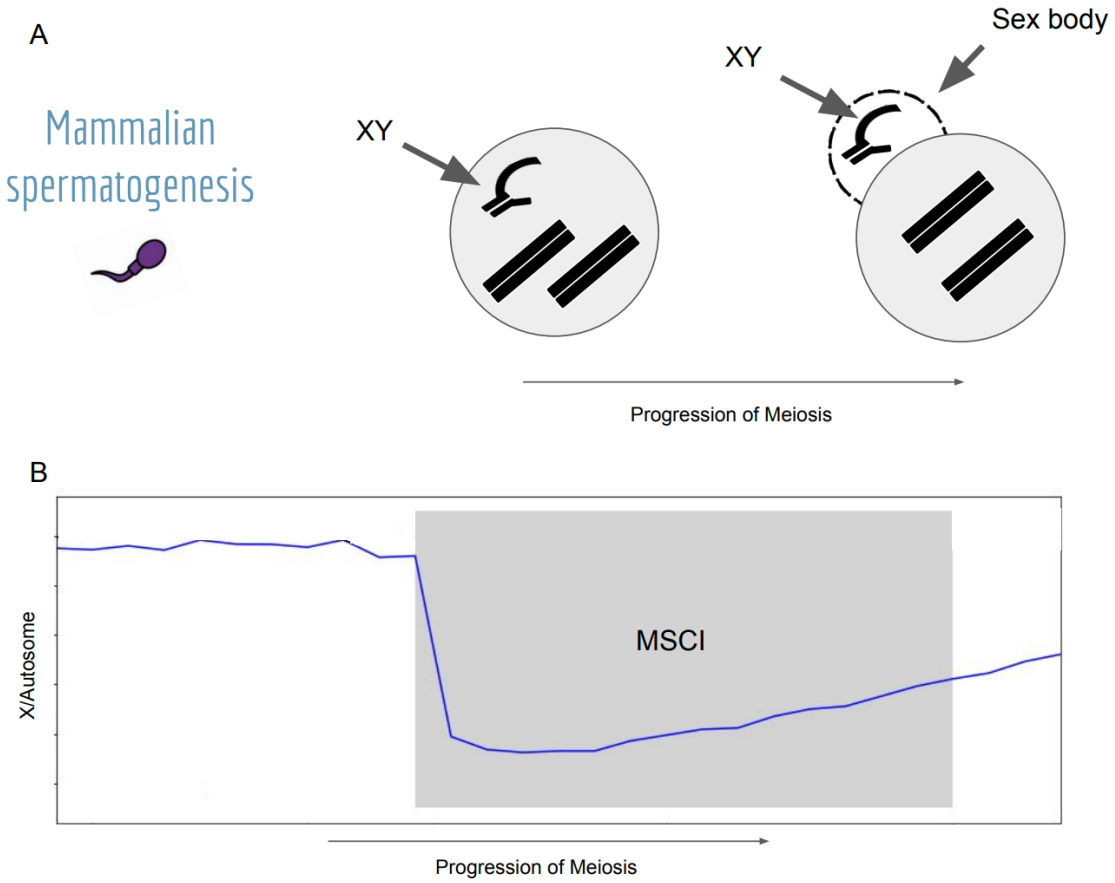


Figure 2: A) A simplified scheme of Meiotic sex-chromosome inactivation in mammalian spermatogenesis, where the XY pair forms a specialized nuclear compartment (sex body). B) An illustration of the X to Autosome expression patterns during meiosis in the presence of MSCI.

Meiosis and its modifications

Although sexual reproduction is near-ubiquitous in eukaryotes, many examples of transitions to asexual reproduction (parthenogenesis) are observed in nature, with a higher occurrence rate in invertebrates compared to vertebrates (38). The different forms of asexual reproduction can be grouped into apomixis (ameiotic reproduction) and automixis (modified meiosis) (39). In apomixis, the individuals produced are genetic clones of the parent; automixis, on the other hand, can produce individuals identical or distinct from the parent, depending on the specific modification to the meiotic pathway and the presence or absence of recombination (40,41). Examples of apomixis in arthropods include two hymenopteran species, the parasitoid wasp and the trichogramma wasp, where pathogenesis by gamete duplication is induced by Wolbachia (41–43). Multiple cases of automixis have also been reported in various arthropods, including the case of the cape honey bees, where worker bees can produce diploid offspring by central fusion automixis (fusion of two non-sister pro-nuclei, Figure 3) (44). In *Daphnia pulex*, a branchiopod crustacean, in a process that has similar outcomes to central fusion, meiosis I is aborted at anaphase I, so the first meiotic division does not take place, but meiosis II proceeds normally (40,44,45). In another form of automixis, terminal fusion (Figure 3), the egg fuses with its sister nucleus from the second meiotic division. This type of automixis has been reported in mayflies, termites, and oribatid mites (40). The different forms of automixis have different consequences on the level of genetic diversity maintained in the species and on its fitness (46).

As meiosis is derived from mitosis (47), many insights about its evolution can be deduced by looking at the genes and mitotic pathways that were re-purposed for meiosis (48); however,

many aspects are eukaryotic innovations specific to meiosis. Those aspects include the pairing between homologous chromosomes and their separation (49), the efficient recombination during pairing, the suppression of the separation between sister-chromatids in the first division, and the absence of the S phase in the second division (47,50). For those aspects and many others, systems with modified meiosis present great opportunities, as they can be used to tackle many of the mechanistic and evolutionary questions surrounding this process. This is achievable by dissecting the differences between normal and modified meiosis and studying the different evolutionary trajectories they entail. Questions that can be tackled using such comparisons include the reasons why some modifications to meiosis are more common than others, why they rise in the first place, and whether there are any prerequisites that make their emergence in some species more likely than in others (40,46).

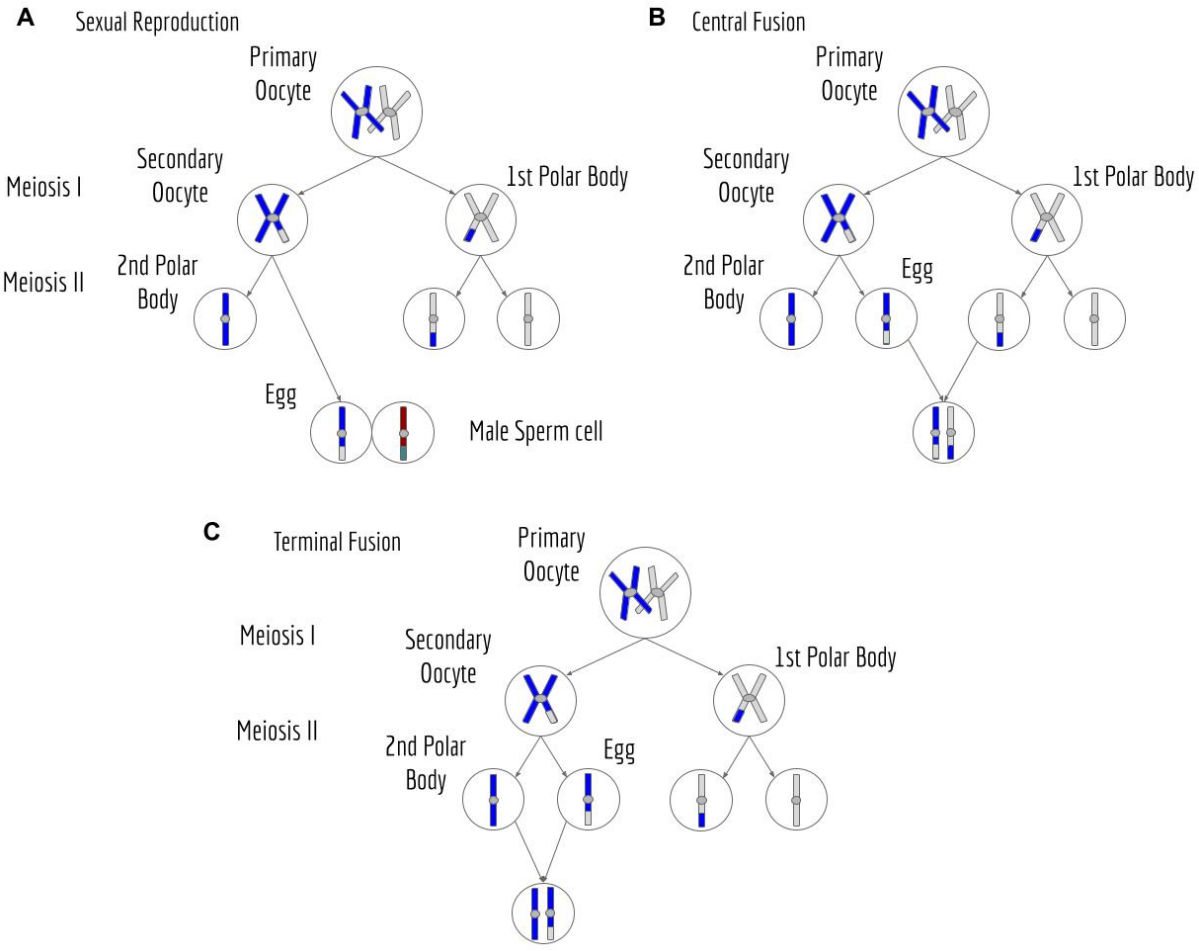


Figure 3: A schematic of regular female meiosis and two of its modifications. A. Regular Female meiosis and sexual reproduction, where the haploid egg from meiosis II fuses with a male sperm cell to restore diploidy. B) Central Fusion Automixis, where the egg fuses with the product of the division of the first polar body to restore diploidy. C) Terminal Fusion Automixis, where the egg fuses with the second polar body to restore diploidy.

Artemia brine shrimp

Artemia brine shrimp live in hundreds of natural and man-made saline waterbodies around the world, in all but cold climate zones. In nature, they disperse to new environments through wind and waterfowl, but in some cases, they are introduced by humans, either to improve salt production or for aquaculture. The genus includes several bisexual species, all diploid, and a large number of parthenogenetic populations, with different levels of ploidy (51). Many species have been described in the literature, either based on sequence variation of genetic markers, reproductive isolation, and/or morphological characters, but it is still hard to give an exact number of *Artemia* species and fully resolve the phylogenetic relationships between them (52,53). *Artemia* can be divided into New and Old world species. The New world species include the invasive north american species of *Artemia franciscana*, and *Artemia persimilis*, found in Argentina and Chile (53). While the new world species are all bisexual, in the Old world, parthenogenetic populations co-occur with the sexual populations of *Artemia salina*, *Artemia sinica*, *Artemia urmiana*, and *Artemia tibetiana*, to name a few, in multiple localities in Eurasia and Australia (54). It has been known for decades that *Artemia* brine shrimp have ZW sex chromosomes (52,54,55); however, at the start of the work presented in this thesis, the nature of these chromosomes and whether they are conserved across the clade was unknown. Given that there are not many ZW species with characterized ZW chromosomes, this system seemed like a great candidate for testing the hypotheses and theories of sex

chromosome evolution, which have been mostly based on studies of XY systems. Chapter 2 describes the development of genomic resources for the clade, and how we used them for investigating sex chromosome evolution in this clade. Chapters 3 and 4 address related questions: how sex is determined at the molecular level in this group, and how the Z-chromosome is regulated in the female germline. Finally, the presence of close sexual and asexual lineages makes *Artemia* an ideal model to study the transitions to asexuality. For a while, the asexual lineages were thought to be ancient, but recent evidence suggests they might have evolved less than 80,000 ago through hybridization events and backcrosses to sexual species (56). In chapters 2 and 5, we search for genomic loci and genes that may control this switch.

References

1. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, et al. Sex Determination: Why So Many Ways of Doing It? PLoS Biol. 2014 Jul 1;12(7):e1001899.
2. Kratochvíl L, Stöck M, Rovatsos M, Bullejos M, Herpin A, Jeffries DL, et al. Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. Philos Trans R Soc B Biol Sci. 2021 Sep 13;376(1833):20200097.
3. Vicoso B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol. 2019 Dec;3(12):1632–41.
4. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 2013 Feb;14(2):113–24.
5. Bull JJ. Evolution of Sex Determining Mechanisms. Benjamin/Cummings Publishing Company, Advanced Book Program; 1983. 344 p.
6. Charlesworth D. When and how do sex-linked regions become sex chromosomes? Evolution. 2021;75(3):569–81.
7. Cheng MK, Disteche CM. A balancing act between the X chromosome and the autosomes. J Biol. 2006;5(1):2.
8. Chandler CH. When and why does sex chromosome dosage compensation evolve? Ann N Y Acad Sci. 2017;1389(1):37–51.
9. Duan J, Larschan EN. Dosage Compensation: How to Be Compensated…Or Not? Curr Biol CB. 2019 Dec 2;29(23):R1229–31.
10. Ercan S. Mechanisms of X Chromosome Dosage Compensation. J Genomics. 2015 Jan 1;3:1–19.
11. McQueen HA, McBride D, Miele G, Bird AP, Clinton M. Dosage compensation in birds. Curr Biol CB. 2001 Feb 20;11(4):253–7.
12. Zimmer F, Harrison PW, Dessimoz C, Mank JE. Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome. Genome Biol Evol. 2016 Apr 1;8(4):1233–42.
13. Lister NC, Milton AM, Patel HR, Waters SA, Hanrahan BJ, McIntyre KL, et al. Incomplete transcriptional dosage compensation of chicken and platypus sex chromosomes is balanced by post-transcriptional compensation. Proc Natl Acad Sci. 2024 Aug 6;121(32):e2322360121.
14. Webster TH, Vannan A, Pinto BJ, Denbrock G, Morales M, Dolby GA, et al. Lack of Dosage Balance and Incomplete Dosage Compensation in the ZZ/ZW Gila Monster (Heloderma suspectum) Revealed by De Novo Genome Assembly. Genome Biol Evol. 2024 Feb 6;16(3):evae018.
15. Kalita AI, Marois E, Kozielska M, Weissing FJ, Jaouen E, Möckel MM, et al. The sex-specific factor SOA controls dosage compensation in Anopheles mosquitoes. Nature. 2023 Nov;623(7985):175–82.
16. Marin R, Cortez D, Lamanna F, Pradeepa MM, Leushkin E, Julien P, et al. Convergent origination of a Drosophila-like dosage compensation mechanism in a reptile lineage. Genome Res. 2017 Dec;27(12):1974–87.
17. Kiss AE, Venkatasubramani AV, Pathirana D, Krause S, Sparr AC, Hasenauer J, et al. Processivity and specificity of histone acetylation by the male-specific lethal complex. Nucleic Acids Res. 2024 May 22;52(9):4889–905.
18. Cerase A, Pintacuda G, Tattermusch A, Avner P. Xist localization and function: new insights from multiple levels. Genome Biol. 2015 Aug 15;16(1):166.
19. Tenorio M, Cruz-Ruiz S, Encarnación-Guevara S, Hernández M, Corona-Gomez JA, Sheccid-Santiago F, et al. MAYEX is an old long noncoding RNA recruited for X chromosome dosage compensation in a reptile. Science. 2024 Sep 20;385(6715):1347–54.
20. Ota R, Hayashi M, Morita S, Miura H, Kobayashi S. Absence of X-chromosome dosage compensation in the primordial germ cells of *Drosophila* embryos. Sci Rep. 2021 Mar 1;11:4890.
21. Mattimoe T, Payer B. The compleX balancing act of controlling X-chromosome dosage and how it impacts mammalian germline development. Biochem J. 2023 Apr 26;480(8):521–37.

22. Sangrithi MN, Royo H, Mahadevaiah SK, Ojarikre O, Bhaw L, Sesay A, et al. Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. Dev Cell. 2017 Feb 6;40(3):289-301.e3.
23. Wei KHC, Chatla K, Bachtrog D. Single-cell RNA-seq of *Drosophila miranda* testis reveals the evolution and trajectory of germline sex chromosome regulation. PLOS Biol. 2024 Apr 30;22(4):e3002605.
24. Turner JMA. Meiotic sex chromosome inactivation. Development. 2007 May 15;134(10):1823–31.
25. Abe H, Alavattam KG, Hu YC, Pang Q, Andreassen PR, Hegde RS, et al. The initiation of meiotic sex chromosome inactivation sequesters DNA damage signaling from autosomes in mouse spermatogenesis. Curr Biol CB. 2020 Feb 3;30(3):408-420.e5.
26. Alavattam KG, Maezawa S, Andreassen PR, Namekawa SH. Meiotic sex chromosome inactivation and the XY body: a phase separation hypothesis. Cell Mol Life Sci CMLS. 2021 Dec 31;79(1):18.
27. Handel MA. The XY body: a specialized meiotic chromatin domain. Exp Cell Res. 2004 May 15;296(1):57–63.
28. Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, et al. The molecular evolution of spermatogenesis across mammals. Nature. 2023 Jan;613(7943):308–16.
29. Maine EM. Meiotic silencing in *Caenorhabditis elegans*. In: Jeon KW, editor. International Review of Cell and Molecular Biology [Internet]. Academic Press; 2010 [cited 2025 Jan 17]. p. 91–134. (International Review of Cell and Molecular Biology; vol. 282). Available from: https://www.sciencedirect.com/science/article/pii/S1937644810820027
30. Guioli S, Lovell-Badge R, Turner JMA. Error-Prone ZW Pairing and No Evidence for Meiotic Sex Chromosome Inactivation in the Chicken Germ Line. Hassold TJ, editor. PLoS Genet. 2012 Mar 8;8(3):e1002560.
31. Mahadevaraju S, Fear JM, Akeju M, Galletta BJ, Pinheiro MMLS, Avelino CC, et al. Dynamic sex chromosome expression in *Drosophila* male germ cells. Nat Commun. 2021 Feb 9;12(1):892.
32. Turner JMA, Mahadevaiah SK, Ellis PJI, Mitchell MJ, Burgoyne PS. Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. Dev Cell. 2006 Apr;10(4):521–9.
33. Larson BJ, Van MV, Nakayama T, Engebrecht J. Plasticity in the Meiotic Epigenetic Landscape of Sex Chromosomes in Caenorhabditis Species. Genetics. 2016 Aug;203(4):1641–58.
34. Anderson J, Henikoff S, Ahmad K. Chromosome-specific maturation of the epigenome in the *Drosophila* male germline. eLife [Internet]. 2023 Nov 17 [cited 2024 Jan 29];12. Available from: https://elifesciences.org/reviewed-preprints/89373
35. Avelino CC, Mendonca CA, Goldstein G, Bernardo CA, Vibranovski MD. Meiotic Sex Chromosome Inactivation: conservation across the *Drosophila* genus [Internet]. bioRxiv; 2024 [cited 2025 Jan 17]. p. 2024.12.04.626848. Available from: https://www.biorxiv.org/content/10.1101/2024.12.04.626848v1
36. Schoenmakers S, Wassenaar E, Hoogerbrugge JW, Laven JSE, Grootegoed JA, Baarends WM. Female Meiotic Sex Chromosome Inactivation in Chicken. Lee JT, editor. PLoS Genet. 2009 May 22;5(5):e1000466.
37. Traut W, Schubert V, Daliková M, Marec F, Sahara K. Activity and inactivity of moth sex chromosomes in somatic and meiotic cells. Chromosoma. 2019 Dec;128(4):533–45.
38. van der Kooi CJ, Matthey-Doret C, Schwander T. Evolution and comparative ecology of parthenogenesis in haplodiploid arthropods. Evol Lett. 2017 Nov 9;1(6):304–16.
39. Archetti M. Recombination and loss of complementation: a more than two-fold cost for parthenogenesis. J Evol Biol. 2004;17(5):1084–97.
40. Engelstädter J. Asexual but Not Clonal: Evolutionary Processes in Automictic Populations. Genetics. 2017 Jun;206(2):993–1009.
41. Jaron KS, Bast J, Nowell RW, Ranallo-Benavidez TR, Robinson-Rechavi M, Schwander T. Genomic Features of Parthenogenetic Animals. J Hered. 2020 Sep 28;112(1):19–33.

42. Pannebakker BA, Pijnacker LP, Zwaan BJ, Beukeboom LW. Cytology of Wolbachia-induced parthenogenesis in Leptopilina clavipes (Hymenoptera: Figitidae). Genome. 2004 Apr;47(2):299–303.
43. Tsutsui Y, Maeto K, Hamaguchi K, Isaki Y, Takami Y, Naito T, et al. Apomictic parthenogenesis in a parasitoid wasp Meteorus pulchricornis, uncommon in the haplodiploid order Hymenoptera. Bull Entomol Res. 2014 Jun;104(3):307–13.
44. Webster MT. Population Genomics: How Do Cape Honey Bees Do Without Sex? Curr Biol. 2020 Jul 20;30(14):R820–1.
45. Hiruta C, Nishida C, Tochinai S. Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol. 2010 Nov;18(7):833–40.
46. Lenormand T, Engelstädter J, Johnston SE, Wijnker E, Haag CR. Evolutionary mysteries in meiosis. Philos Trans R Soc B Biol Sci [Internet]. 2016 Oct 19 [cited 2024 Feb 14]; Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0001
47. Wilkins AS, Holliday R. The evolution of meiosis from mitosis. Genetics. 2009 Jan;181(1):3–12.
48. Bernstein H, Bernstein C. Evolutionary Origin of Recombination during Meiosis. BioScience. 2010 Aug 1;60(7):498–505.
49. McKee BD. Homologous pairing and chromosome dynamics in meiosis and mitosis. Biochim Biophys Acta BBA - Gene Struct Expr. 2004 Mar 15;1677(1):165–80.
50. Ohkura H. Meiosis: An Overview of Key Differences from Mitosis. Cold Spring Harb Perspect Biol. 2015 May;7(5):a015859.
51. Asem A, Eimanifar A, Rastegar-Pouyani N, Hontoria F, Vos SD, Stappen GV, et al. An overview on the nomenclatural and phylogenetic problems of native Asian brine shrimps of the genus Artemia Leach, 1819 (Crustacea, Anostraca). ZooKeys. 2020 Oct 13;902:1–15.
52. Bowen ST. THE GENETICS OF ARTEMIA SALINA. III. EFFECTS OF X-IRRADIATION AND OF FREEZING UPON CYSTS. Biol Bull. 1963 Dec;125(3):431–40.
53. Parraguez M, Gajardo G, Beardmore JA. The New World Artemia species A. franciscana and A. persimilis are highly differentiated for chromosome size and heterochromatin content. Hereditas. 2009 May;146(2):93–103.
54. Maccari M, Gómez A, Hontoria F, Amat F. Functional rare males in diploid parthenogenetic Artemia. J Evol Biol. 2013 Sep 1;26(9):1934–48.
55. Huylmans AK, Toups MA, Macon A, Gammerdinger WJ, Vicoso B. Sex-Biased Gene Expression and Dosage Compensation on the Artemia franciscana Z-Chromosome. Mank J, editor. Genome Biol Evol. 2019 Apr 1;11(4):1033–44.
56. Rode NO, Jabbour-Zahab R, Boyer L, Flaven É, Hontoria F, Van Stappen G, et al. The Origin of Asexual Brine Shrimps. Am Nat. 2022 Aug;200(2):E52–76.

CHAPTER 2

ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp

Published

Chapter 2: General Abstract

To study several aspects of reproduction in our model organism *Artemia*, we produced two chromosome-level assemblies: one for the Asian species of *Artemia sinica*, and the other for the American species of *Artemia franciscana*. We used the *Artemia sinica* genome to show the conservation of the sex chromosomes across all the *Artemia species*, to shed light on the genetic origin of rare males, and to search for loci that co-segregate with parthenogenesis. We have also performed synteny analysis between all the chromosomes of *Artemia sinica* and *Artemia franciscana* and showed that they are highly syntenic. We used a combination of coverage and Fst estimates to characterize the extent of differentiation across the different regions of the Z chromosome. Our detection of various Z/W gametologs in the younger strata made it possible to estimate their relative age using the rate of synonymous divergence (dS). The ancient stratum, on the other hand, seems to be largely degenerated, and we are uncertain if the gametologs we detect in it were ancestrally there. However, its conservation across the clade suggests a minimum age of 40 million years.

Then this chapter contains two publications:

Elkrewi M*, Khauratovich U*, Toups MA*, Bett VK, Mrnjavac A, Macon A, Fraisse C, Sax L, Huylmans AK, Hontoria F, Vicoso B. ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp. Genetics. 2022 Oct 1;222(2):iyac123.

In this publication, I was responsible for the genome assembly, the identification of the W candidates in the two species using a K-mer substraction approach, and the expression analysis. I was also in charge of repeating parts of the *Artemia parthenogenetica* (Aibi lake) rare male and female heterozygosity analysis. I also generated all the figures for the analyses I performed.

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

I was responsible for the conceptualization of the assembly pipeline, and the execution was a joint effort with Vincent Bett. Additionally, I performed the synteny analysis, and the detection and comparing of the strata across the two species.

Chapter 2.1: ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp

Marwan Elkrewi[1#], Uladzislava Khauratovich[12#], Melissa A. Toups[13#], Vincent Kiplangat Bett[1], Andrea Mrnjavac[1], Ariana Macon[1], Christelle Fraisse[14], Luca Sax[15], Ann Kathrin Huylmans[16], Francisco Hontoria[7], Beatriz Vicoso[1@]

[1] Institute of Science and Technology Austria, Klosterneuburg, Austria
[2] Department of Chromosome Biology, Max Perutz Labs, University of Vienna, Vienna BioCenter, Vienna, Austria
[3] Department of Life and Environmental Sciences, Faculty of Science and Technology, Bournemouth University, Dorset, UK
[4] CNRS, Univ. Lille, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France
[5] Lewis and Clark College, Portland, OR, USA
[6] Institute of Organismic and Molecular Evolution, Johannes Guttenberg Universität Mainz, Mainz, Castellón, Germany
[7] Instituto de Acuicultura de Torre de la Sal (IATS-CSIC), Spain
[@] bvicoso@ist.ac.at

[#] These authors contributed equally to this work.

Abstract
Eurasian brine shrimp (genus *Artemia*) have closely related sexual and asexual lineages of parthenogenetic females, which produce rare males at low frequencies. Although they are known to have ZW chromosomes, these are not well characterized, and it is unclear whether they are shared across the clade. Furthermore, the underlying genetic architecture of the transmission of asexuality, which can occur when rare males mate with closely related sexual females, is not well understood. We produced a chromosome-level assembly for the sexual Eurasian species *A. sinica* and characterized in detail the pair of sex chromosomes of this species. We combined this new assembly with short-read genomic data for the sexual species *A. sp. Kazakhstan* and several asexual lineages of *A. parthenogenetica,* allowing us to perform a first in-depth characterization of sex-chromosome evolution across the genus. We identified a small differentiated region of the ZW pair that is shared by all sexual and asexual lineages, supporting the shared ancestry of the sex chromosomes. We also inferred that recombination suppression has spread to larger sections of the chromosome independently in the American and Eurasian lineages. Finally, we took advantage of a rare male, which we backcrossed to sexual females, to explore the genetic basis of asexuality. Our results suggest that parthenogenesis is likely partly controlled by a locus on the Z chromosome, highlighting the interplay between sex determination and asexuality.

Introduction

The diversity of reproductive and sex-determining systems has long puzzled evolutionary biologists (Bachtrog *et al.* 2014; Pennell *et al.* 2018; Picard *et al.* 2021). When separate sexes are present, the development of males and females can be controlled by environmental factors or through the presence of sex-determining loci (Beukeboom and Perrin 2014; Bachtrog *et al.* 2014). These sex determining loci are typically carried by specialized “sex chromosomes”, such as the X and Y chromosomes of mammals. Sex chromosomes initially arise from standard pairs of autosomes, but can progressively stop recombining over much of their length, ultimately resulting in genetic and morphological differentiation (Charlesworth *et al.* 2005; Wright *et al.* 2016). Each segment of the sex chromosome pair that stopped

recombining at a given timepoint is referred to as a "stratum", and strata of different ages are often found on the same pair of sex chromosomes (Lahn and Page 1999; Handley *et al.* 2004). The Y chromosome stops recombining altogether after XY recombination suppression and eventually degenerates, i.e., it accumulates deleterious mutations and can lose many or even all of its genes (Bachtrog 2013). This gene loss leads to dosage deficits in males, since many X-linked genes become single-copy. Mechanisms of dosage compensation often target the X-chromosome and regulate its expression, thereby reestablishing optimal dosage balance of genes across the genome (Charlesworth 1978; Straub and Becker 2007; Vicoso and Bachtrog 2009; Disteche 2016). Alternatively, both the silencing of Y-linked genes and compensation of X-linked genes may arise concurrently as a result of runaway regulatory divergence that sets up and reinforces the predominance of X over Y expression (Lenormand *et al.* 2020; Lenormand and Roze 2022). Much of our understanding of these processes has come from studying the ancient XY systems of traditional model organisms such as mice and fruit flies. Despite the recent characterization of young sex chromosomes in many nonmodel species (Charlesworth 2019), many questions remain about the earlier stages of sex-chromosome divergence. For example, what molecular mechanisms and selective pressures drive the initial loss of recombination between sex chromosomes (Ponnikas *et al.* 2018)? Similarly, female-heterogametic species (i.e., females are ZW, males are ZZ) have remained relatively understudied, as they are not found in any of the main model organisms. While parallels exist between the evolution of XY and ZW pairs, such as the progressive loss of recombination and subsequent degradation of the Y/W-chromosomes (Ellegren 2011; Vicoso *et al.* 2013; Zhou *et al.* 2014; Picard *et al.* 2018; Sigeman *et al.* 2021), some aspects of their evolution seem to differ. In particular, dosage compensation of Z-chromosomes is often limited to a few dosage-sensitive genes (i.e., it works gene-by-gene, as opposed to the chromosome-wide mechanisms found in many XY species, (Mank 2013; Rovatsos and Kratochvíl 2021)). These discrepancies may have to do with systematic differences in selection and mutation between males and females (Vicoso and Bachtrog 2009; Ellegren 2011; Mullon *et al.* 2015), or may simply be a coincidence due to the few ZW systems characterized in detail at the molecular level (Rovatsos and Kratochvíl 2021).

Although the prevalence of sexual reproduction suggests that it offers long-term advantages, asexual lineages are found in many clades and successfully inhabit a variety of ecological niches (Toman and Flegr 2018). Transitions from sexual to asexual reproduction are frequent (Neiman et al., 2014), and can involve a diversity of mechanisms that disrupt meiosis, such as novel mutations, hybridization of closely related lineages, and polyploidization (Neiman *et al.* 2014). Asexuality can evolve from any ancestral sex-determining system, including in species with differentiated sex chromosomes, (e.g., Schwander and Crespi 2009; Jaquiéry *et al.* 2014; Mignerot *et al.* 2019), and understanding the mechanisms underlying these transitions has been a key goal of the field.

In many asexual lineages, males are occasionally produced, and can fertilize closely related sexual females, which then give rise to new asexual lineages (“contagious parthenogenesis”). These crosses have facilitated the use of traditional genetic approaches for understanding the genetic architecture of asexuality (Jaquiéry *et al.* 2014). Transitions from sexual to asexual reproduction have primarily been studied in animal species where both sexual reproduction and parthenogenesis were ancestrally part of the life cycle, either in the form of cyclical parthenogenesis or haplodiploidy (Neiman *et al.* 2014). In this case, the loss of sexual reproduction and consequent obligatory parthenogenesis is often controlled by one or only a few loci (Lynch *et al.* 2008; Sandrock and Vorburger 2011; Eads *et al.* 2012; Jaquiéry *et al.* 2014; Aumer *et al.* 2017; Yagound *et al.* 2020). In the pea aphid, the locus controlling asexuality is found on the X-chromosome (Jaquiéry *et al.* 2014), and a locus of large effect on parthenogenesis was also found on the UV sex chromosome pair of brown algae *Ectocarpus* (Mignerot *et al.* 2019), raising interesting questions about the interplay between the ancestral sex-determining system and contagious parthenogenesis. One direct link between the two phenomena is that when asexuals are derived from an ancestral XX/XY or haplodiploid sex-determination systems, rare males can be formed through the loss of an X-chromosome (Kampfraath *et al.* 2020) or through accidental production of haploid individuals during

automixis (Sandrock and Vorburger 2011). Less is known about the creation of rare males when the ancestral sex-determination system was female-heterogamety. More generally, it is unclear if sex chromosomes are a prime spot for the location of genes regulating asexual reproduction, since very few transitions have been characterized in organisms with sex chromosomes.

Brine shrimp of the genus *Artemia* have both asexual and sexual species (Abatzopoulos 2018), as well as ZW sex chromosomes with putative ancient and recent strata (Bowen 1963; De Vos *et al.* 2013; Accioly *et al.* 2015; Huylmans *et al.* 2019), making them an ideal model for addressing many of these questions. While all American species are sexual, the Eurasian clade consists of a few sexual species (including *A. sinica, A. sp. Kazakhstan* and *A. urmiana*) and of various asexual lineages (collectively refered to as *A. parthenogenetica*, and further referred to by their location of origin) (Van Stappen 2002; Maccari *et al.* 2013b). Asexuals vary in ploidy, but only diploid lineages are considered here (Maccari *et al.* 2013b). Originally thought of as ancient, these lineages turned out to have arisen recently through hybridization between asexual lineages and individuals from, or closely related to, *A. sp. Kazakhstan* (Baxevanis *et al.* 2006; Maccari *et al.* 2013b; Rode *et al.* 2022). In *Artemia*, such contagious parthenogenesis can occur through the production of rare males by asexual lineages, which can fertilize closely related sexual females (Maccari *et al.* 2013a; Abatzopoulos 2018). Furthermore, asexual females can mate with males of sexual species and produce a minority of offspring sexually (Boyer *et al.* 2021). The ZW pair of *Artemia* has been mostly studied in the American species *A. franciscana* (Bowen 1963; Parraguez *et al.* 2009; De Vos *et al.* 2013; Accioly *et al.* 2015). Both a small differentiated region and a non-recombining but largely undifferentiated region were detected, making it an interesting system to understand the first steps leading to ZW divergence (Huylmans *et al.* 2019). Gene expression in the differentiated region appears to be fully balanced between males and females, but there was limited power to detect changes due to the fragmented nature of the genome (Huylmans *et al.* 2019). Eurasian lineages also carry a ZW pair (Haag *et al.* 2017), but whether the same chromosome is used for sex determination accross the clade in not known. Because *A. parthenogenetica* reproduce through central fusion automixis (Nougué *et al.* 2015), a modified form of meiosis, which allows for loss of heterozygosity when recombination between chromosomes occurs, rare recombination events between the Z and W (which replace part of the W with its Z-linked homologous region) can lead to the creation of rare males (Nougué *et al.* 2015; Boyer *et al.* 2022). Finally, the genetic mechanisms behind asexuality, and whether the sex chromosomes play any further role in its evolution, have not yet been explored in detail.

Here, we develop several genomic resources for *Artemia* lineages, including the first chromosome-level assembly for the *Artemia* genus (*A. sinica*), as well as short-read genomic data for *A. sp. Kazakhstan* and several lineages of *A. parthenogenetica* (see Sup. Fig. 1 for a phylogeny of the lineages). Using these data, we are able to provide an in-depth characterization of sex-chromosome evolution across the genus, including identifying an ancient region shared with the American species *A. franciscana*. Finally, we find evidence that asexuality is likely partly controlled by a locus on the Z chromosome - a first in a ZW sex chromosome system.

Results

1. The ZW pair is shared by American and Eurasian *Artemia*

Two genome assemblies and a high density linkage map are currently available for the American *A. franciscana* (Jo *et al.* 2021a; Han *et al.* 2021; De Vos *et al.* 2021), but resources for the Eurasian clade are more limited, with only an *A. sp. Kazakhstan* draft genome assembly recently described in Boyer et al. (2022). The median dS (the number of synonymous substitutions per synonymous site) between the two clades is ~0.2. We assembled a male genome of *A. sinica* using PacBio long reads (~30x) and Hi-C Illumina reads (1.5*e12 reads), yielding 1213 scaffolds with an N50 of 67.19 Mb (Sup. Fig. 2, Sup. Table 1) and a total length of 1.7Gb; 85% of the sequences get assigned to one of the 21 largest scaffolds (which corresponds to the expected number of chromosomes, Sainz-Escudero *et al.* 2021). The

strong diagonal in the heatmap of the Hi-C contact matrix (Sup. Fig. 3) supports the high quality of our assembly, as does our BUSCO score of 91.8%. This chromosome-level assembly represents an improvement over existing *Artemia* genomes, which have N50 values of 27 to 112Kb, and BUSCO scores of 68.3% to 86.9% (Jo *et al.* 2021a; De Vos *et al.* 2021; Boyer *et al.* 2022; Sup. Fig. 4).

Our earlier analysis of female and male genomic coverage in *A. franciscana* had uncovered a small region of reduced female coverage, consistent with full differentiation of the Z and W chromosomes (Huylmans *et al.* 2019). To investigate whether ZW differentiation was also present in *A. sinica*, we first estimated male and female coverage along each chromosome. Consistent with *A. franciscana*, only a small genomic region on chromosome 1 had decreased female/male coverage (Fig. 1A, Sup. Fig. 5 for all chromosomes), showing that chromosome 1 is the Z chromosome. To check for homology with the *A. franciscana* differentiated region, we mapped the scaffolds from the *A. franciscana* genome of (Jo *et al.* 2021a) to the new *A. sinica* assembly based on their shared gene content, and plotted the coverage values that we had previously estimated (Huylmans *et al.* 2019) based on the *A. sinica* coordinates. Fig. 1A shows that the two differentiated regions largely overlap, supporting the ancestry of the pair of sex chromosomes; we name this shared region stratum 0 (S0). In the *A. franciscana* linkage map (Han *et al.* 2021), LG6 was identified as the sex chromosome. To further verify the homology between the ZW pairs of the two species, we mapped the genetic markers used by Han et al. (2021) to our *A. sinica* assembly. As expected, the vast majority of LG6 markers for which we could infer a location mapped to our chromosome 1 (Sup. Fig. 6). We also produced an assembly based on *A. sinica* female long PacBio reads, which contains a substantial amount of scaffolds with excessive female coverage, consistent with W-linkage (Sup. Fig. 7).

2. Convergent loss of ZW recombination

To identify parts of the sex chromosomes that no longer recombine, but are still similar enough that W-derived reads still map to the Z, we used previously published RNA-seq data for *A. sinica* (Huylmans *et al.* 2021), obtained from 10 males and 10 females, to estimate F_{ST}, a measure of genetic differentiation, between the two sexes. Genetic variants found exclusively on the W increase the level of female-male differentiation, and young non-recombining regions can be detected through their high male:female F_{ST} (Palmer *et al.* 2019; Vicoso 2019; Gammerdinger *et al.* 2020). Such an increase in male:female F_{ST} is not expected for the highly differentiated S0, since W-derived reads do not map to this part of the Z-chromosome. Fig. 1B shows that a large region (~52 Mb) has F_{ST} values systematically above the 95th-percentile of autosomes, consistent with recent loss of recombination in *A. sinica*. We call this region Stratum 1 (S1), but further divide it into S1a, which shows localized drops in female:male coverage (gray shaded regions in Fig. 1A), and S1b, for which no coverage differences are observed (Fig. 1A). The distal end of S1a has reduced female:male coverage in *A. franciscana*, and an F_{ST} analysis in this species yielded increased male:female FST from ~60 to 85MB (see Sup. Fig. 8), showing that at least part of this region has also stopped recombining in the American lineage.

Given the substantial distance between the Eurasian and American lineages, we hypothesized that the loss of recombination in S1 had occurred independently in the two clades. To test this, we used a k-mer-based pipeline combining male and female DNA and RNA short reads (Elkrewi *et al.* 2021) to identify putative W-derived transcripts. This yielded 402 transcripts in *A. franciscana* and 319 in *A. sinica*. Of those that mapped to the genome, 180 out of 306 (59%) *A. sinica* transcripts and 168 out of 355 (47%) *A. franciscana* transcripts mapped to chromosome 1 (Z) of *A. sinica*, a higher proportion than the overall 7% of genes that map to this chromosome, confirming the validity of the approach (since we expect many W-linked genes to have a close homolog on the Z). Few of these candidate W genes mapped to the putative ancestral sex-linked region (16 W-linked genes in *A. sinica*, compared to 84 Z-linked genes, and 7 versus 91 in *A. franciscana*, Sup. Table 2), consistent with substantial degeneration of this part of the W-chromosome. To find genes present on the W-chromosomes of both species, we selected reciprocal best hits between the two sets of W

candidates. All 15 candidates that were found in both species mapped to the putative S1a region. We made phylogenetic trees using each pair of homologous W-genes and their Z-linked homologs (obtained from a male-only transcriptome assembly, and all mapping to the S1a region), to infer whether these genes were W-linked before the split of the two clades. The closest homolog in the transcriptome of the distantly related fairy shrimp *Branchinecta lindahli* (Schwentner *et al.* 2018) was used as an outgroup sequence, when one could be detected. Fig. 1C shows the resulting phylogenetic trees for two of the shared W-linked genes and their Z-homologs, while phylogenies for all candidates are provided in Sup. Fig. 9. In every case, ZW homologs clustered by species rather than by chromosome, confirming that loss of recombination occurred independently and convergently for this region in the American and Eurasian lineages.

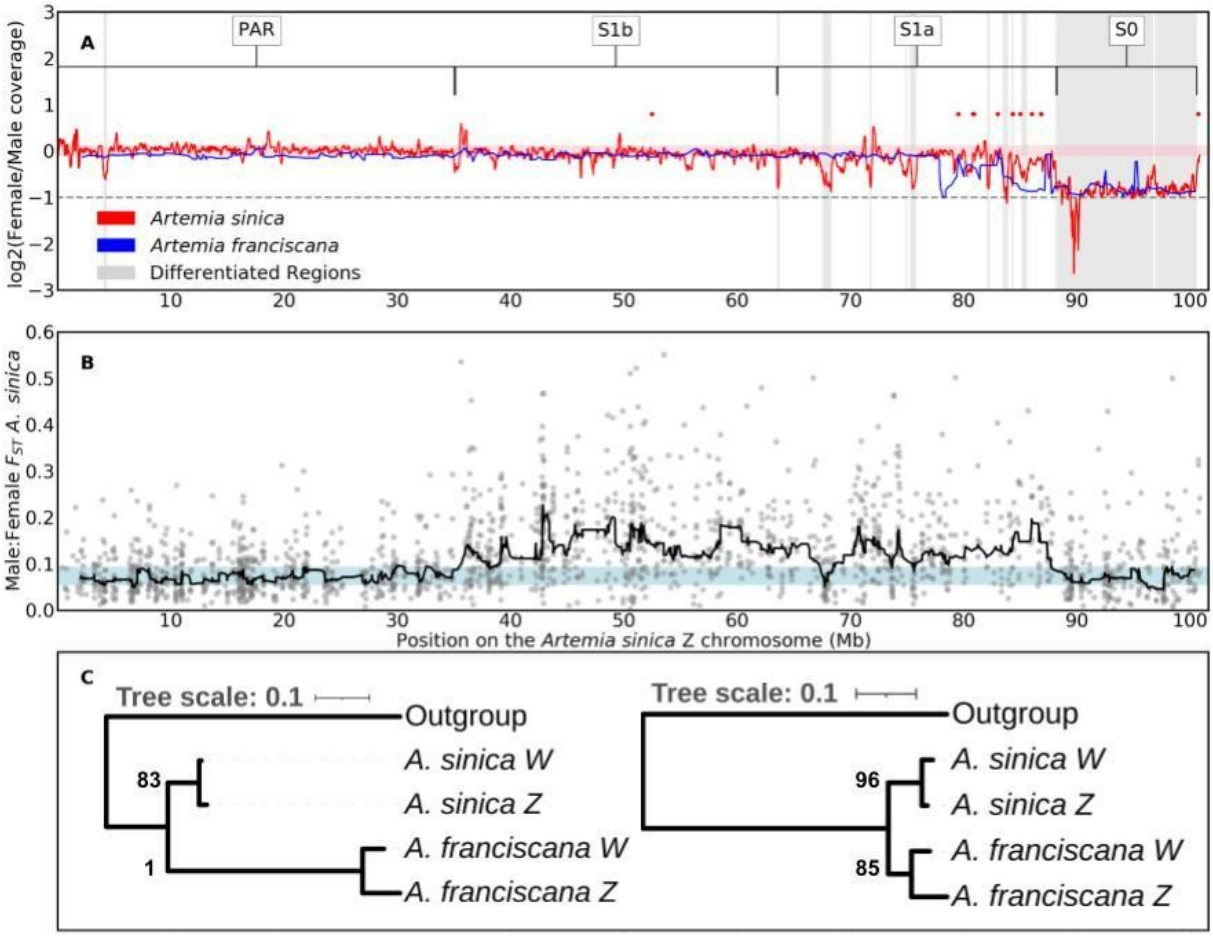


Fig.1: A shared sex-linked region on the ZW pair. A. Patterns of female/male coverage in *A. franciscana* and *A. sinica*. A rolling median of 20 windows (of 10KB each) is shown in red for *A. sinica*, and a rolling median of 5 scaffolds is shown in blue for *A. franciscana*. The light pink shaded area highlights the region between the 5th and 95th-percentiles of the rolling median of coverage for autosomes of *A. sinica*. The horizontal gray line at -1 signifies the expected twofold reduction in coverage in females compared to males of fully differentiated regions. The inferred differentiated regions of *A. sinica* are highlighted in gray, and the putative strata are defined above, along with the putative pseudoautosomal region (PAR). The red dots are the locations of the W-candidates shared between *A. sinica* and *A. franciscana*. B. Male:female F_{ST} along the putative chromosome Z. The dots are F_{ST} calculated for bins of 1000 nucleotides, and the black line is the rolling median computed in sliding windows of 30 consecutive 1000 nucleotide bins. The light blue shaded area highlights the region between the 5th and 95th-percentiles of the rolling median of F_{ST} for autosomes. C. Phylogenetic trees for two examples of the W-candidates shared between *A. sinica* and *A. franciscana* (red dots in panel A) and their putative Z homologs. *Branchinecta lindahli* is used as the outgroup.

3. Dosage compensation of the Z-specific region

Many female-heterogametic species lack a chromosome-wide mechanism of dosage compensation, and investigating the few cases that have it may shed light on the difference between ZW and XY systems. Earlier work suggested that the Z-specific region of *A. franciscana* was compensated (Huylmans et al. 2019), but misidentification of genes in the sex-linked region (as the genome was fragmented) could have hidden differences between chromosomes. We repeated this analysis using RNA-seq data from thorax, head, and gonad of *A. sinica* (Huylmans et al. 2021). We first assembled a male transcriptome from all pooled male reads available for this species (to avoid hybrid assemblies of Z and W homologs, see Sup. Fig. 10 for a BUSCO assessment), mapped it to the male genome assembly, and estimated expression for each sample (in transcripts per million, TPM). In somatic tissues, the female:male ratio is similar for the autosomal genes and S0 genes ($p = 0.2$ and $p = 0.6$ in heads and thoraces, Bonferroni-corrected Wilcoxon tests, Fig. 2B and 2C), confirming that dosage compensation is active in this clade. A significant shift towards male-biased expression can be observed for the S0 in gonads ($p = 0.0007$, Bonferroni-corrected Wilcoxon test, Fig. 2D). A table with nominal p-values for each comparison is provided in Sup. Table 3.

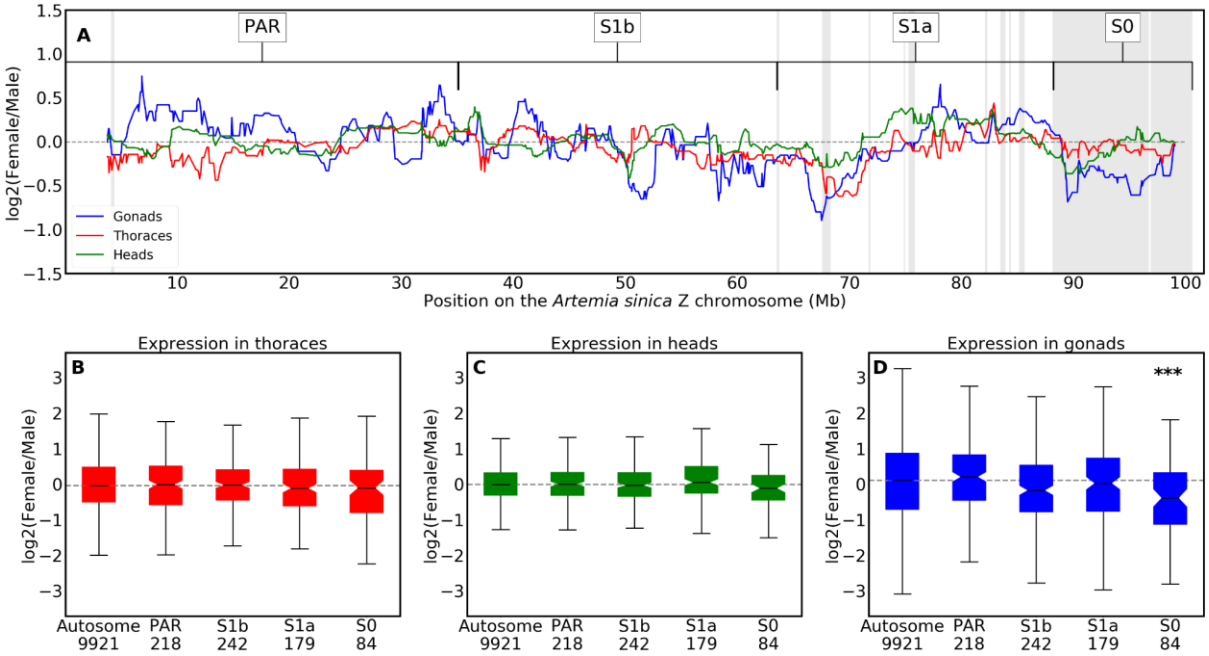


Figure 2. Dosage compensation of the Z-chromosome. Panel A: The log-transformed ratio of female to male expression along the Z chromosome in heads, gonads and thoraces (computed as the rolling median in sliding windows of 30 consecutive genes). Gray areas represent the differentiated regions identified in the coverage analysis, and the putative strata are denoted above, along with the putative pseudoautosomal region (PAR).The dashed horizontal black line is at zero. Panels B-D: The distribution of log-transformed ratio of female to male expression for the autosomes and the different regions of the Z chromosome in thoraces (A), heads (B) and gonads (C). The number of genes in each of the different regions is indicated underneath the x-axis labels. A wilcoxon rank sum test was used to assess the significance of the difference between the expression of the autosomes and the different regions of the Z chromosome, with a Bonferroni correction for the four comparisons performed in each tissue. *** denotes a p-value <=0.001.

4. The sex chromosomes of asexual females and the genetic origin of rare males

In order to characterize the ZW pair of asexual females, we first obtained a draft genome assembly of the closely related sexual species *A. sp. Kazakhstan* from illumina short reads (Sup. Table 4), and estimated genomic coverage using two female and two male samples of this species. The genomic scaffolds were mapped to the *A.sinica* genome based on their gene content, and median coverages of male and female *A. sp. Kazakhstan* individuals were plotted along the *A. sinica* Z chromosome using a sliding window of 10 scaffolds (green and yellow lines in Fig. 3A). As expected, an approximately two-fold drop in female coverage was observed in a similar region to that found in *A. sinica* (marked by gray shading), whereas the male harbored high genomic coverage throughout the chromosome, consistent with the presence of the same pair of sex chromosomes in this lineage (a similar pattern was observed in *A. urmiana*, Sup. Fig. 11). We used the *A. sp. Kazakhstan* draft genome to map genomic reads derived from three closely related asexual females (one from the Lake Urmiana-derived population, and two from a population derived from Aibi Lake cysts). In every case, the patterns of coverage were very similar to those of the *A. sp. Kazakhstan* sexual female, confirming that asexual females carry the same pair of ZW chromosomes.

Diploid *A. parthenogenetica* likely reproduce through central fusion automixis, a modified form of meiosis that preserves heterozygosity in the genome except at distal ends of chromosomes when recombination has occurred (Nougué *et al.* 2015). Boyer *et al.* (2022) recently showed that *Artemia* rare males can be produced by ZW recombination events at variable locations near the sex-determining locus. We obtained a rare male from an *A. parthenogenetica* line from Aibi Lake (which we use in the next section to explore the transmission of asexuality). To test whether it arose through ZW recombination or other chromosomal changes, we first compared patterns of genomic coverage to those of females.

No reduced coverage was observed along the Z-chromosome, arguing against the loss of a sex chromosome. We further called Single Nucleotide Polymorphisms (SNPs) in the rare male and in its sister (marked as A. par. Aibi lake 2 in Fig 3A), and estimated the proportion of heterozygous SNPs present in the asexual female that were lost in the rare male. Loss of heterozygosity was detected throughout the distal half of the Z-chromosome (Fig. 3B and Sup. Fig. 12), confirming that a large part of the W was replaced by its Z homologous region. A substantial loss of heterozygosity was also found at the beginning of chr 13, and smaller regions of decreased heterozygosity may be present at the ends of several chromosomes (Sup. Fig. 10). Taken together, these results support central fusion automixis as the mode of reproduction of *A. parthenogenetica*, and rare ZW recombination as the source of the Aibi Lake rare male (Nougué *et al.* 2015; Boyer *et al.* 2022).

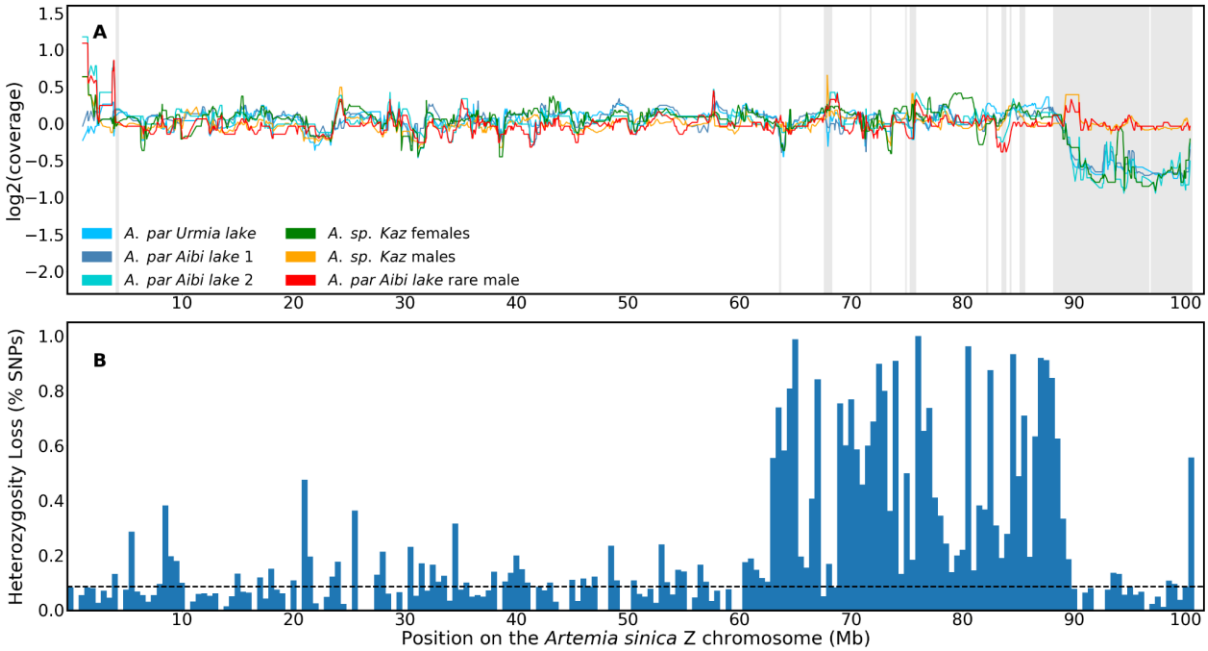


Fig. 3: The sex chromosomes of sexual and asexual individuals. A. Coverage patterns in *A. sp. Kazakhstan* male and female samples, in three asexual females, and in a rare male derived from an asexual lineage from Aibi Lake. Areas shaded in gray represent the differentiated regions of the *A. sinica* ZW pair. B. The fraction of SNPs that lost heterozygosity on the rare male Z chromosome relative to its asexual sister in bins of 500Kb. The dotted line represents the average loss of heterozygosity for autosomes.

5. The Z chromosome likely contributes to the transmission of asexuality

In order to find possible loci responsible for the spread of asexuality in brine shrimp, we crossed the rare male described in the previous section and a sexual female from *A. sp. Kazakhstan* (Sup. Fig. 13). This produced 22 asexual females and 24 males in the F1; one additional female died without producing offspring asexually. The presence of asexual females in the F1 shows that the locus controlling asexuality in this lineage works in a dominant manner, unlike what was first observed in (Maccari *et al.* 2014), but consistent with the recent experiments of Boyer *et al.* (2021). The fact that almost all females produced offspring without mating further suggests that the locus was likely present on both copies of the genome of the original rare male. We then backcrossed 12 males from the F1, which should only carry one copy of the locus/loci controlling asexuality, with females from an *A. sp. Kazakhstan* inbred line (of these only 6 yielded progeny). The resulting F2 generation consisted of 84 (~45%) males, 5 (~3%) asexual females, and 96 (~52%) females that did not produce asexually 133 days after the crosses were set up (44 individuals died before sexing was possible and are not included in the counts; see counts for individual crosses in Sup. Table 5). We presume that most of these are sexual females for our analyses, but some could have reproduced asexually had the experiment been continued longer.

We produced whole-genome resequencing data for the 5 F2 asexual females and, as a control, 10 F2 putatively sexual females. These were first pooled into an asexual pool and a putatively sexual pool, and we used Popoolation2 to compute F_{ST} between these two pools of females. While a few small peaks of F_{ST} are found on the autosomes (Fig 4a), the strongest signal comes from the Z chromosome (70 to 90MB, Fig 4b). It should be noted that it is possible that high F_{ST} extends into the differentiated part of the Z, but that we do not have sufficient coverage to detect it. Consistent with this, when no subsetting of reads is performed to equalize coverage across individuals (see methods), increased F_{ST} is observed throughout the distal end of the Z (Sup. Fig 14). We further predicted that loci underlying asexuality should have been inherited from the original rare male by all the F2 asexual females, but not by (all) control females. To test this, we mapped all DNA samples individually to the *A. sp. Kazakhstan* genome. We also mapped the original rare male and its *A. parthenogenetica* sister, and two *A. sp. Kazakhstan* individuals, in order to select SNPs that were alternatively fixed between the two lineages. We used these informative SNPs to reestimate F_{ST} between F2 asexual and control females, and to infer which genomic regions were inherited from the rare male by each of the F2 individuals. Sup. Fig. 15 shows that we recover a region of high F_{ST} on the Z chromosome, and that all asexuals carry genetic material from the rare male in this region, as expected if it controls asexuality. In total, only 17 scaffolds with an assigned location on the *A. sinica* genome show ancestry patterns consistent with an asexuality locus (i.e., they show evidence of *A. parthenogenetica* ancestry in all asexual females, but not in all control females). Eleven are on the Z chromosome (versus 1 expected, p=1.3e-20 with a Chi-square test), and correspond to the region of high F_{ST}, providing further support for a role of the Z chromosome in the transmission of asexuality. None of the other minor peaks of F_{ST} are in regions with ancestry patterns consistent with asexuality loci (Sup. Fig. 14), although chromosome 16 contains 3 such loci (versus 0.9 expected, $p<0.01$ with a chi-square test). The annotation of genes located in the Z-linked candidate locus did not yield any obvious candidates (the annotation for all transcripts is provided as a Supplementary Dataset).

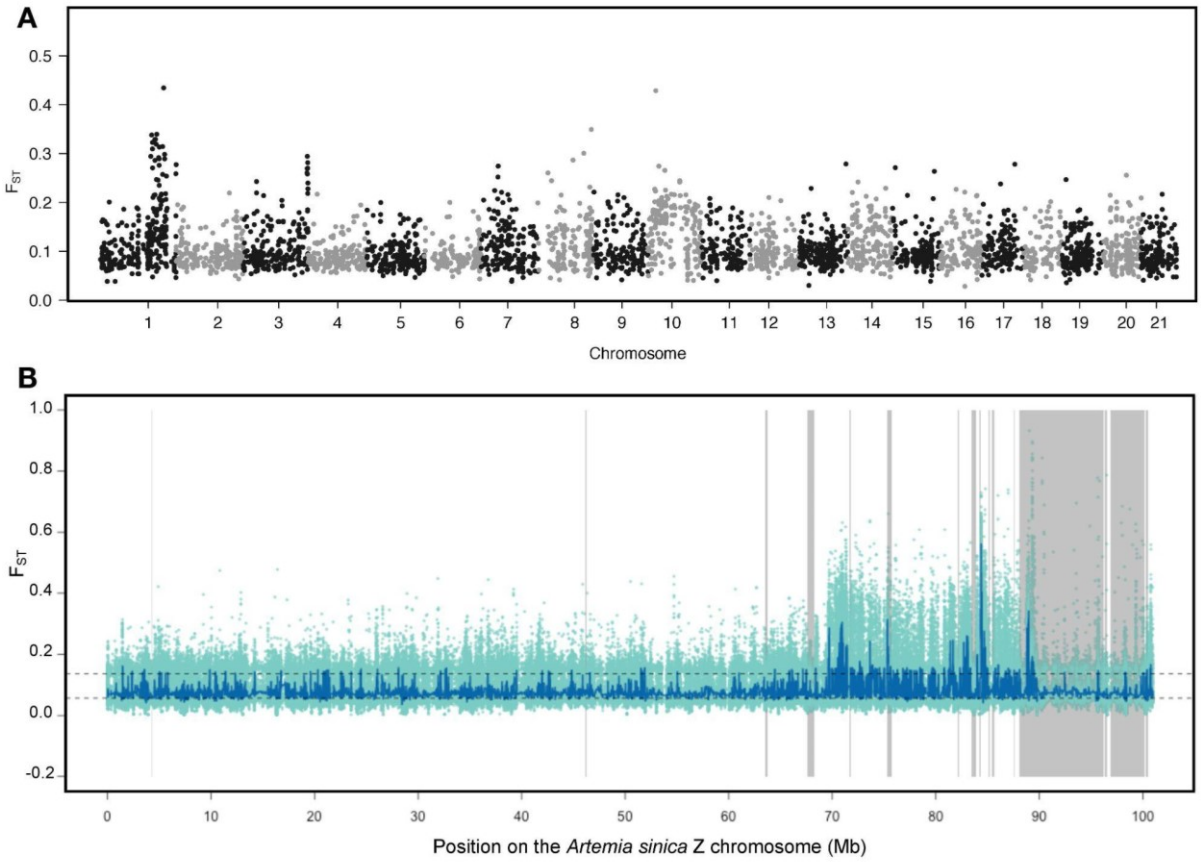


Fig 4. Elevated F_{ST} between sexual and asexual females localizes to the Z chromosome. A. Manhattan plot of F_{ST} estimated for 1Kb sliding windows between asexual and sexual females across the genome. B. F_{ST} across chromosome 1. F_{ST} is shown for individual SNPs in light green, and the blue line shows the rolling median for 101 SNPs. Areas shaded in gray represent the differentiated regions of the *A. sinica* ZW pair.

Discussion

The potential of *Artemia* brine shrimp as models for ZW chromosome evolution and comparative genomics in general was until recently hampered by a lack of genomic resources. The publication of two genomes for the American *A. franciscana* has already shed new light on how these charismatic organisms survive in their extreme environments (Jo *et al.* 2021a; De Vos *et al.* 2021), but no information on sex linkage was provided, and the lack of a close outgroup sequence (other than the distant *Daphnia*) made comparative analyses difficult. A draft genome was recently described for *A. sp. Kazakhstan* (Boyer *et al.* 2022), but there was limited power to assign scaffolds to the sex chromosomes or autosomes.

Here, we obtain the first chromosome-level assembly in the clade for the Eurasian brine shrimp *A. sinica*, and characterize in detail the differentiated and undifferentiated regions of the ZW pair. By combining these results with those of a preliminary analysis in *A. franciscana* (Huylmans *et al.* 2019), we confirmed the putative evolutionary model for the ZW pair, with an ancient well-differentiated region that stopped recombining in the ancestor of the two lineages, and more recent “strata” arising in each lineage independently. The independent loss of recombination in American and Eurasian species provides a unique opportunity to investigate convergent changes that occur in early sex-chromosome evolution. In agreement with previous findings in *A. franciscana*, *A. sinica* males and females have similar somatic expression patterns of Z-linked genes in the differentiated region which strongly supports the presence of a mechanism of dosage compensation in this group. A significant male-bias in expression was found for S0 genes in the gonads. Such differences in the gonad have been found even in animals with well-characterized chromosome-wide mechanisms of dosage

compensation, such as *Drosophila* (Meiklejohn et al. 2011) and silkworm (Huylmans *et al.* 2017). While compensation mechanisms may be absent or less active in the gonad (Meiklejohn *et al.* 2011) , differences could also result from the unusual regulation of the sex chromosomes in the germline (Argyridou and Parsch 2018), where they are often inactivated or downregulated (Vibranovski *et al.* 2009). Currently, no tractable lab model exists for the early evolution of ZW chromosomes, and Z-chromosome dosage compensation is only understood in detail for the silkworm (Walters and Hardcastle 2011; Kiuchi *et al.* 2014; Katsuma *et al.* 2019; Rosin *et al.* 2022), making this an outstanding new model clade for investigating these topics.

Finally, we obtained several putative W-genes in both species using a k-mer based analysis. Few of them mapped to the ancestral part of the W chromosome: only ~20% of the Z-linked genes in this region have a W-homolog, suggesting that much of the ancestral gene content has been lost. All of the genes for which a W-homolog could be found in both *A. sinica* and *A. franciscana* mapped to younger strata of the ZW pair, and appear to have become W-linked independently in the two lineages. If the ancestral sex-determination mechanism is still shared by the two species, this may suggest that the primary signal for sex determination is a dosage-dependent gene on the Z rather than a dominant female-determining gene on the W. However, it is also possible that the sex-determining gene is only expressed early in development, and was missed by our analysis of adult tissues. Future studies of sex-specific expression throughout the life cycle and complete assemblies of the W chromosome of the two species will be necessary to shed light on sex determination in this group.

The proximity of *A. sinica* to the *A. sp. Kazakhstan* group, which contains both sexual and asexual populations, also allowed to characterize sex-linked sequences in this group. First, we found that the sex chromosome pair is shared by all populations. We further confirmed that rare males in this group can be produced through the replacement of the W-specific region with its Z-counterpart, showing that the same mechanism is used for rare male production in *A. parthenogenetica* isolates from widely differing geographic origins (China in this study, Iran and France in Boyer et al., 2022). Finally, our backcrossing experiment points to a role of the sex chromosome pair in the spread of asexuality through rare males.

It should be noted that this experiment has several drawbacks. First, it is difficult to phenotype females as sexual or asexual, as the timing at which asexuals produce their first brood can vary (although they typically do so within 30 days of hatching, Amat *et al.* 2007). Furthermore, hybrid incompatibilities may stop females from producing viable offspring even if they carry the alleles encoding asexuality. The fact that only ~5% of females were asexual in the F2 suggests that either the trait is polygenic, and/or that we are mistakenly classifying asexuals as putative sexuals. Finally, we only obtained a small number of backcrossed asexual females, which limits the power to infer causal loci.

Despite these drawbacks, the Z chromosome showing the strongest signal of differentiation between asexual and control females is intriguing, and in line with results in the pea aphid, which carries the asexuality locus on the X chromosome. In species where asexuality is triggered by an endoparasite such as Wolbachia, the acquisition of asexuality is thought to be driven by the transmission advantage gained by the female-transmitted parasite (since asexual reproduction leads to an all-female progeny). It is possible that an asexuality gene found on a Z chromosome similarly benefits from a transmission advantage. If rare males always arise through the replacement of the W-specific region with its Z homolog region, a Z-linked asexuality locus will be homozygous and therefore transmitted to all daughters in the F1 (and to all sons). More detailed studies of transmission of asexuality in this group and others with ZW and XY sex chromosomes will in the future shed light on the relationship between sex determination and the rise and spread of asexual reproduction under various sex-determining mechanisms.

Materials and methods

Sampling and DNA extractions

Cysts from *A. sinica* (originally from Tanggu salterns, PR China), *A. sp. Kazakhstan* (originally from an unknown location in Kazakhstan), and two lineages of *A. parthenogenetica* (from Lake Aibi (PR China) and from Lake Urmia (Iran)) were obtained from the Instituto de Acuicultura de Torre de la Sal (C.S.I.C.) *Artemia* cyst collection in Spain, as described in (Huylmans *et al.* 2021). Cysts were hatched under 30 g/L salinity and grown to adulthood under 60 g/L salinity. Some of these F0 individuals were used directly for DNA extractions with the Qiagen DNeasy Blood & Tissue kit. We also set up iso-female lines in *A. sinica* and *A. sp. Kazakhstan*, and subjected them to 6 generations of sib-sib mating to reduce the amount of heterozygosity. Male and female individuals from *A. sinica* and *A. sp. Kazakhstan* inbred iso-female lines were used individually for DNA extractions with Qiagen DNeasy Blood & Tissue kit. Furthermore, 20 males and 17 females of *A. sinica* (also from the inbred iso-female line) were pooled and high molecular weight DNA was extracted using the Qiagen Genomic-tip 20/G kit.

DNA short and long read sequencing

PacBio libraries were prepared and sequenced at the Vienna Biocenter Sequencing facility for the male and female *A. sinica* high molecular weight DNA. All other DNA samples were used for Illumina paired-end sequencing. Libraries were prepared and sequenced at the Vienna Biocenter Sequencing Facility. Finally, 1 male was frozen and provided to the sequencing facility for Hi-C library preparation and Illumina sequencing on a NovaSeq machine. The final list of samples, as well as the parts of the analysis that they were used in, are listed in Sup. Table 6.

Genome assemblies

The male PacBio reads were assembled using two different genome assemblers: Flye (v.2.7.1, Kolmogorov *et al.* 2019) and Miniasm (0.3-r179, minimap2 2.18-r1028-dirty was used for mapping and the consensus was generated using Racon v1.4.22)(Li 2016; Vaser *et al.* 2017). The Flye assembly was polished using male *Artemia sinica* short genomic reads (trimmed with the Trimmomatic package, Bolger *et al.* 2014), and the Miniasm assembly was polished using the same male short reads using the wtpoa-cns tool from wtdbg2 (version 2.5, Ruan and Li 2020). The two assemblies were then merged using quickmerge (version 0.3, Chakraborty *et al.* 2016) with the Miniasm assembly as the query and the Flye assembly as the reference. The resulting assembly was purged based on the male pacbio read depth to remove duplicates and contig overlaps using the purge_dups program (version 1.2.5, Guan *et al.* 2020).

To scaffold the assembly into pseudo-chromosomes, the PCR duplicates were first removed from the Hi-C data using the clumpify.sh script from the bbmap package (Bushnell 2014), and the Hi-C reads were then mapped to the genome assembly using the Arima mapping pipeline with MAPQ 5 (Arima Genomics 2021) and then scaffolded using the YaHS tool (pre-release of version 1.1, Zhou *et al.* 2022). The contact maps were visualized and manually edited on Juicebox (version 1.11.08, Robinson *et al.* 2018) to generate the final chromosome-level assembly.

The female *Artemia sinica* genome was assembled from female PacBio reads using Flye (version 2.7.1), and it was not polished to avoid collapsing the Z and the W scaffolds. To identify putative W scaffolds, short genomic reads from two *A. sinica* males and two females were mapped to the female assembly using Bowtie2 (Langmead and Salzberg 2012). We then counted how many male and female reads mapped to each scaffold, after filtering for alignments without mismatches (by selecting only mapped reads with the CIGAR string “NM:i:0”). The female-specific Kmers inferred to obtain W-specific transcripts (section ”Identification of candidate W-genes with Kmer analysis” below) were similarly mapped to each scaffold with Bowtie2 and counted. Scaffolds which had more than 5 female-specific

Kmers, and more perfect matches in females than in males (male/(male+female)<=0.3) were considered candidates W-derived scaffolds, and are highlighted in orange in Sup. Fig. 7.

The *Artemia sp. Kazakhstan* genome was assembled from two male short read libraries with Megahit (v1.1.4, Li *et al.* 2015) and then scaffolded using SOAPdenovo-fusion (SOAPdenovo2 version 2.04, Luo *et al.* 2012).

BUSCO (version 5.2.2, Manni *et al.* 2021) was used to assess the completeness of the genomes generated in this study and the two previously published *Artemia franciscana* genomes in the genome mode with the arthropoda dataset (arthropoda_odb10).

Estimation of genomic coverage

The short genomic reads were mapped to the genome using bowtie2 (version 2.4.4, Langmead and Salzberg 2012). The uniquely mapped reads were then extracted from the output sam files using (grep -vw "XS:i"). SOAP.coverage (version 2.7.7, Luo *et al.* 2012) was then used to calculate the coverage for each library either using 10000 bp windows (*A. sinica*) or per scaffold (other species).

Mapping of *A. franciscana* markers to the A. sinica genome

The sequences of the *A. franciscana* SLAF markers were obtained from Han *et al.* (2021), and the left and right pairs of each marker were mapped to the *Artemia sinica* male genome separately using pblat (Wang and Kong 2019). Only the mapping location with the largest match score was kept for each transcript.

Mapping of the *A. franciscana* and *A. sp. Kazakhstan* genomes to the new *A. sinica* assembly

We aligned the *A. sinica* published transcriptome (Huylmans *et al.* 2021) to both the *A. franciscana* and to the *A. sp Kazakhstan* genomic scaffolds using blat (Standalone BLAT v. 36x2, Kent 2002). For each transcript, we kept only the mapping location with the highest score in each genome (using the customized script 1-besthitblat.pl). When multiple transcripts overlapped by more than 20bps on the genome, only the transcript with the highest mapping score was kept (2-redremov_blat_v2.pl). We then used the location of the transcripts on the *Artemia sinica* genome to infer the location of the *A. franciscana* and *A. sp. Kazakhstan* scaffolds based on the transcripts they carried (AssignScaffoldLocation.pl). This script uses a majority rule to assign each scaffold to a chromosome, and then the mean location of genes on that scaffold to infer its final coordinate on the chromosome. All scripts are available on our git page.

F_{ST} between male and female populations

RNA-seq reads from 10 pooled *A. sinica* males and 10 pooled *A. sinica* females (from Huylmans et al, 2021), sampled from head, thorax and gonads, were pooled by sex and mapped separately to the male *A.sinica* reference genome using STAR (Dobin *et al.* 2013) with default parameters.

The resulting alignments with MAPQ score lower than 20 were filtered out and the remaining alignments were sorted using samtools view and sort functions (Li *et al.* 2009). Next, a pileup file of male and female alignments was produced using the samtools-mpileup function. Finally, we used scripts from the Popoolation2 package (Kofler *et al.* 2011) to calculate F_{ST}. The mpileup file was reformatted with the Popoolation2 mpileup2sync.pl script, and the resulting synchronized file was used as an input for fst-sliding.pl script. F_{ST} between male and female populations was calculated for windows of 1000 nucleotides, using the fst-sliding.pl script with following options --suppress-noninformative --min-count 3 --min-coverage 10 --max-coverage 200 --min-covered-fraction 0.5 --window-size 1000 --step-size 1000 --pool-size 10.

We applied the same pipeline to estimate male:female FST using head and gonad RNA-seq samples obtained from 10 males and 10 females of *A. franciscana* (from Huylmans et *al.*, 2019). The resulting F_{ST} values were plotted based on the inferred location of the genomic

scaffolds along the *A. franciscana* chromosomes (section “Mapping of the *A. franciscana* and *A. sp. Kazakhstan* genomes to the new *A. sinica* assembly”).

Strata identification

ZW strata were identified based on the *A. sinica* coverage and F_{ST} analyses. First, we detected differentiated regions as any region where the Log2(female/male coverage) dropped below (median(autosomal windows) - 0.5)) for 10 consecutive 10KB windows; each differentiated region was extended along the chromosome as long as Log2(female/male coverage) did not rise above that threshold for 10 consecutive windows (regions shaded in gray in Fig. 1). The two largest differentiated regions were nearly adjacent on the distal end of chromosome 1, and the whole region encompassing them was classified as S0 (no genes were found in the small undifferentiated region between them, such that including it in the S0 did not affect downstream analyses). We used a similar approach to detect regions of increased male:female F_{ST}. In this case, only sparse information along the chromosome was obtained (as RNA-seq only provides SNPs for genic regions), and many 1kb bins were empty. We selected only informative bins, and inferred an F_{ST} rolling median for 30 bins at a time (the median coordinate for the bins was similarly used as the coordinate for the resulting window). High F_{ST} regions were called when 10 consecutive rolling windows were above the 95%-percentile of autosomal windows, and these regions were extended along the chromosome until 10 consecutive windows were below this threshold. This yielded three nearly-adjacent high F_{ST} sections (35.3-38.2MB, 38.4-67.3MB, 68.5-87.7MB), and the region encompassing them (35.3-87.7MB) was classified as S1. S1 was further divided into S1a, which showed drops in female:male coverage, and S1b, which did not. The coordinate of the beginning of the first differentiated region within S1 was used as the boundary between them.

Identification of candidate W-genes with Kmer analysis

We used a k-mer based subtraction approach (Elkrewi et al., 2021) based on the tools included in the BBMap package (Bushnell, 2014) on male and female genomic and RNA-seq data from *A. franciscana* and *A. sinica*. The pipeline was applied to each species separately. In *A. sinica,* two male and two female DNA libraries and two whole body RNA-seq replicates for each sex were used (Sup. Tables 6 and 7). In *A. franciscana*, the analysis was performed using one male and one female DNA libraries and pools of two RNA-seq replicates of heads and gonads for each sex, along with one whole body male and female RNA-seq libraries (SRR14598203 and SRR14598204).

First, the shared 31-mers between the female DNA and RNA libraries were identified, and then any k-mers matching male libraries were removed. Female RNA-seq reads containing these female-specific Kmers [with minimum kmer fraction of 0.6 (mkf=0.6)] were retrieved and assembled using Trinity (Grabherr *et al.* 2011), and the perl script from the Trinity package (get_longest_isoform_seq_per_trinity_gene.pl) was used to keep only the longest isoform. The male and female genomic libraries were mapped to the assembled transcripts using Bowtie2 (Langmead and Salzberg 2012 p. 2), and candidates with a sum of female perfect matches <=8 and a ratio of sum-of-females/sum-of-males <=2 were removed. The final set consisted of 402 transcripts in *A. franciscana* and 319 in *A. sinica*.

Mapping of W candidates to the *A. sinica* genome

The *A. sinica* and *A. franciscana* candidate W-derived transcripts were mapped to the *A. sinica* genome assembly with Parallel Blat (Wang and Kong 2019) with a translated query and database, and a minimum match score of 50. Only alignments with match scores above 100 were considered, and the mapping location with the strongest match score was considered for each transcript.

Transcriptome assemblies and expression analysis

The *A. sinica* male transcriptome was assembled from two replicates of male whole body RNA-seq data (Huylmans *et al.* 2021) using Trinity (Grabherr *et al.* 2011) in two different modes: denovo and genome-guided. The two assemblies were concatenated and then filtered

using the tr2aacds.pl script from EvidentialGene (Gilbert 2019). The transcriptome was annotated with the Pannzer annotation server (Törönen *et al.* 2018), and mapped to the *A. sinica* genome using the same procedure as described in section “Mapping of W candidates to the A. sinica genome”.

For the expression analysis, only the first isoform was kept for each gene, and only transcripts longer than 500bp were used in the analysis. The RNA-seq reads from the *A. sinica* heads, gonads, and thoraces of males and females (Huylmans *et al.* 2021) were mapped to the curated transcriptome and gene expression levels (in Transcripts per million, TPM) were obtained using Kallisto (version 0.46.2, Bray *et al.* 2016). Normalization was done using NormalyzerDE (Willforss *et al.* 2019).

Two different *A. franciscana* de novo transcriptome assemblies were made using Trinity. The first using pooled RNA-seq reads from male heads and testes (two replicates each, (Huylmans *et al.* 2019)), and the second using the published whole-body male RNA-seq library (SRR14598203, Jo *et al.* 2021b). The two assemblies were concatenated and then filtered using the tr2aacds.pl script from EvidentialGene, and mapped to the *A. sinica* genome using the same procedure as described in section “Mapping of W candidates to the A. sinica genome”.

Phylogenetic Trees

The W candidates of *A. sinica* and *A. franciscana* were mapped reciprocally to each other using pblat (v. 36x2 with default parameters, Wang and Kong 2019), and reciprocal best hits were considered shared candidates. The W candidates of the two species were further mapped to their respective uncollapsed male transcriptome assemblies (see previous section) with pblat (Wang and Kong 2019) with a translated query and database, and a minimum match score of 50. The transcripts with the highest mapping score to the W candidates were used as the putative Z homologs in their respective species.

The *Branchinecta lindahli* transcriptome (Schwentner *et al.* 2018) was downloaded from the Crustacean Phylogeny dataset on Harvard Dataverse (https://doi.org/10.7910/DVN/SM7DIU). *B. lindahli* homologs of shared W-candidates were obtained by mapping the putative Z homologs of both species to the *B. lindahli* transcriptome using pblat (-minScore=50 -t=dnax -q=dnax) and retrieving the transcript with highest alignments score (using the customized script 2-besthitblat.pl). A transcript was considered a homolog if it mapped to at least one of the putative Z homologs of the two species, and when the two Z homologs mapped to different outgroup sequences, both outgroup sequences were retrieved and used to make two different alignments.

The shared W candidates of *A. sinica* and *A. franciscana*, their Z homologs, and the outgroup sequences were aligned using MAFFT (version v7.487, with the options “mafft --adjustdirection INPUT > OUTPUT”, Katoh *et al.* 2002). The resulting alignments were fed to phylogeny.fr (Dereeper *et al.* 2008), where the alignment was curated using GBLOCKS (Talavera and Castresana 2007), and the phylogenetic tree was constructed using PhyML (Guindon *et al.* 2010). Trees were made only for sequences where the number of overlapping positions after gblocks was longer than 200bp (Sup. Fig. 9). In the four instances where the curated alignment length with the outgroup was shorter than 200bp, we tried aligning the sequences without the outgroup. For the two cases where the resulting alignment length was longer than or equal 200bp, unrooted trees were made (Sup. Fig. 9). The trees were then downloaded in the Newick format and visualized using itol.embl.de (Letunic and Bork 2019).

Heterozygosity analysis in asexual female and rare male

Illumina genomic sequencing was performed on a rare male and its asexual sister (both derived from an Aibi Lake *A. parthenogenetica* lineage), yielding around 115 million paired-end reads with a length of 125 nucleotides for each sample. The reads were quality- and adapter- trimmed with Trimmomatic-0.36 (Bolger *et al.* 2014), and mapped to the draft *Artemia sp. Kazakhstan* genome assembly using STAR v.2.6.0c (Dobin *et al.* 2013) with default settings.

We indexed the reference *A. sp. Kazakhstan* genome using SAMtools v.1.10 (Li *et al.* 2009), called the SNPs from BAM alignments with BCFtools v.1.10.2 (Li *et al.* 2009), then removed indels, filtered for quality of reads over 30 and coverage over 5 and below 100 with VCFtools v.0.1.15 (Danecek *et al.* 2011), and removed multiallelic sites with BCFtools.

We calculated the fraction of SNPs that lost heterozygosity in the rare male DNA in comparison with the asexual sister DNA. It was calculated and visualized in 500kb bins for each chromosome.

Crossing design to identify the asexuality locus

We designed a backcross in order to investigate the loci controlling asexuality (Sup. Fig. 13). An asexual female from Aibi Lake produced a rare male. We crossed this male with an inbred female from the closest related sexual species, *A. sp. Kazakhstan*. This produced asexual females and males in the F1 generation. We then backcrossed 12 males from the F1 to sexual females from from the same inbred line of *A. sp. Kazakhstan*. Of these, 6 crosses produced offspring, yielding a total of 84 males, 5 asexual females, and 96 putatively sexual females (those that did not reproduce asexually for 133 days after the crosses were set up). The 5 asexual females and 10 control females were used individually for DNA extractions with the Qiagen DNeasy Blood & Tissue kit. The control females came from the same crosses (i.e., had the same F2 father and *A. sp. Kazakhstan* mothers) as the asexual females, but were otherwise selected randomly. Illumina short-read sequencing was then performed as described in section “DNA short and long read sequencing”.

Analysis of backcross between the Aibi Lake rare male and *A. sp. Kazakhstan* females

We sequenced 5 asexual females and 10 putatively sexual females from the F2 generation. This resulted in an average of 101 million reads per asexual female and 50 million reads per putatively sexual female. We first used SEQTK v1.2 (https://github.com/lh3/seqtk) to randomly select a subset of reads from each asexual sample to match the number of reads of the smallest sample (to avoid biasing allele estimates towards high-coverage individuals). We removed adaptors and trimmed reads using Trimmomatic v0.39 (Bolger *et al.* 2014). We then aligned the resulting paired-end reads to the genome using Bowtie2 v2.4.4 (Langmead and Salzberg 2012). SAM files were converted to BAM files and sorted in Samtools v.1.13 (Li *et al.* 2009).

For our pooled analyses, we merged BAM files into a pooled asexual BAM file and a pooled putatively-sexual BAM, and created a mpileup file in Samtools v.1.13. We then used Popoolation2 (Kofler *et al.* 2011) to call F_{ST} for both individual SNPs and in 1kb windows. We used F_{ST} computed for 1kb windows to visualize F_{ST} across the genome in a Manhattan plot in the R package qqman (Turner 2018). We computed rolling medians in sliding windows of 101 consecutive SNPs on each linkage group using the rollmedian function from the package zoo (Zeileis and Grothendieck 2005) in R v.4.0.3. To identify regions of elevated F_{ST} on individual chromosomes, we computed 95% confidence intervals by sampling rolling medians of 101 consecutive SNPs across the genome 1000 times.

For our individual-based analyses, we similarly used SEQTK v1.2 to randomly select a subset of reads from each asexual sample to match the highest coverage found in an F2 control female (to avoid biases caused by the much larger number of reads obtained for the F2 asexuals than for the controls). We then mapped reads from all F2 individuals to the *A. sp. Kazakhstan* genome using BWA mem v0.7.17 (Li and Durbin 2009) with default parameters. DNA reads from the rare male and its *A. parthenogenetica* sister, and from two *A. sp. Kazakhstan* individuals, were also subsetted and mapped. The resulting BAM alignments were sorted with samtools v1.14 (Li *et al.* 2009), and used to call SNPs with the mpileup function of BCFtools v1.14 (Li 2011). The VCF file was filtered with VCFtools v0.1.17 (Danecek *et al.* 2011) for minimum and maximum depth (4 and 50), minimum quality score (30) and minimum minor allele frequency (0.1). Only SNPs for which the two *A. sp. Kazakhstan* had a genotype of 0/0, and the two *A. parthenogenetica* individuals 1/1, were kept for further analyses. We computed F_{ST} between the F2 asexual and control females using the function --weir-fst-pop in VCFtools for 10kb windows. We then inferred ancestry of each genomic

scaffold in every sample (i.e., whether they were homozygous for the *A. sp. Kazakhstan* haplotype, or carried a copy of the *A. parthenogenetica* haplotype as well) using the customized script Chromopaint.pl (available on our git page). The *A. sp. Kazakhstan* genomic scaffolds were assigned to a location on the *A. sinica* genome as before. Scaffolds with more than 10 informative SNPs, and >80% 0/1 or 1/1 SNPs were considered to be heterozygous for the *A. sp. Kazakhstan* and *A. parthenogenetica* haplotypes, whereas scaffolds with >80% 0/0 were considered to have only *A. sp. Kazakhstan* ancestry (only 5 to 9% of scaffolds fell in between and could not be classified in each individual).

Data availability statement

All genomic reads generated for this study are available at the NCBI short reads archive under Bioproject number PRJNA848277. The pipelines used to analyze the data are at https://git.ist.ac.at/bvicoso/zsexasex2021, and important processed data files such as the new A. sinica genome assembly are provided at https://doi.org/10.15479/AT:ISTA:11653.

Acknowledgements

We thank the Vicoso group for comments on the manuscript, and the ISTA Scientific computing team and the Vienna Biocenter Sequencing facility for technical support. This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 715257), and by the Austrian Science Foundation (FWF SFB F88-10).

References

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

Accioly I. V., I. M. C. da Cunha, J. C. M. Tavares, and W. F. Molina, 2015 CHROMOSOME BANDING IN CRUSTACEA. I. KARYOTYPE, Ag-NORs, C BANDING AND TREATMENT WITH EcoRI, PstI and KpnI RESTRICTION ENDONUCLEASES IN *Artemia franciscana*. Biota Amaz. Biote Amaz. Biota Amazon. Amaz. Biota.

Amat F., F. Hontoria, J. C. Navarro, N. Vieira, and G. Mura, 2007 Biodiversity loss in the genus *Artemia* in the Western Mediterranean Region. Limnética 26: 387–404.

Argyridou E., and J. Parsch, 2018 Regulation of the X Chromosome in the Germline and Soma of *Drosophila melanogaster* Males. Genes 9: 242. https://doi.org/10.3390/genes9050242

Arima Genomics, 2021 *mapping_pipeline*. Arima Genomics, Inc. https://github.com/ArimaGenomics/mapping_pipeline

Aumer D., M. H. Allsopp, H. M. G. Lattorff, R. F. A. Moritz, and A. Jarosch-Perlow, 2017 Thelytoky in Cape honeybees (Apis mellifera capensis) is controlled by a single recessive locus**.** Apidologie volume 48, pages 401–410.

Bachtrog D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14: 113–124. https://doi.org/10.1038/nrg3366

Bachtrog D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, *et al.*, 2014 Sex Determination: Why So Many Ways of Doing It? PLOS Biol. 12: e1001899. https://doi.org/10.1371/journal.pbio.1001899

Baxevanis A. D., I. Kappas, and T. J. Abatzopoulos, 2006 Molecular phylogenetics and asexuality in the brine shrimp *Artemia*. Mol. Phylogenet. Evol. 40: 724–738. https://doi.org/10.1016/j.ympev.2006.04.010

Beukeboom L. W., and N. Perrin, 2014 *The Evolution of Sex Determination*. Oxford University Press, Oxford.

Bolger A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowen S. T., 1963 The Genetics of *Artemia salina*. III. Effects of X-Irradiation and of Freezing upon Cysts. Biol. Bull. 125: 431–440. https://doi.org/10.2307/1539357

Boyer L., R. Jabbour-Zahab, M. Mosna, C. R. Haag, and T. Lenormand, 2021 Not so clonal asexuals: Unraveling the secret sex life of *Artemia parthenogenetica*. Evol. Lett. 5: 164–174. https://doi.org/10.1002/evl3.216

Boyer L., R. Jabbour-Zahab, P. Joncour, S. Glémin, C. R. Haag, *et al.*, 2022 Asexual male production by ZW recombination in *Artemia parthenogenetica*. bioRxiv 2022.04.01.486774; doi: https://doi.org/10.1101/2022.04.01.486774

Bray N. L., H. Pimentel, P. Melsted, and L. Pachter, 2016 Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34: 525–527. https://doi.org/10.1038/nbt.3519

Bushnell B., 2014 BBMap: A Fast, Accurate, Splice-Aware Aligner**.** No. LBNL-7065E. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA.

Chakraborty M., J. G. Baldwin-Brown, A. D. Long, and J. J. Emerson, 2016 Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 44: e147. https://doi.org/10.1093/nar/gkw654

Charlesworth B., 1978 Model for evolution of Y chromosomes and dosage compensation. Proc. Natl. Acad. Sci. 75: 5618–5622. 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: 118–128. https://doi.org/10.1038/sj.hdy.6800697

Charlesworth D., 2019 Young sex chromosomes in plants and animals. New Phytol. 224: 1095–1107. https://doi.org/10.1111/nph.16002

Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Vos S., P. Bossier, G. Van Stappen, I. Vercauteren, P. Sorgeloos, *et al.*, 2013 A first AFLP-

based genetic linkage map for brine shrimp *Artemia franciscana* and its application in mapping the sex locus. PloS One 8: e57585. https://doi.org/10.1371/journal.pone.0057585
De Vos S., S. Rombauts, L. Coussement, W. Dermauw, M. Vuylsteke, *et al.*, 2021 The genome of the extremophile *Artemia* provides insight into strategies to cope with extreme environments. BMC Genomics 22: 635. https://doi.org/10.1186/s12864-021-07937-z
Dereeper A., V. Guignon, G. Blanc, S. Audic, S. Buffet, *et al.*, 2008 Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36: W465–W469. https://doi.org/10.1093/nar/gkn180
Disteche C. M., 2016 Dosage compensation of the sex chromosomes and autosomes. Semin. Cell Dev. Biol. 56: 9–18. https://doi.org/10.1016/j.semcdb.2016.04.013
Dobin A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635
Eads B. D., D. Tsuchiya, J. Andrews, M. Lynch, and M. E. Zolan, 2012 The spread of a transposon insertion in Rec8 is associated with obligate asexuality in Daphnia. Proc. Natl. Acad. Sci. 109: 858–863. https://doi.org/10.1073/pnas.1119667109
Elkrewi M., M. A. Moldovan, M. A. L. Picard, and B. Vicoso, 2021 Schistosome W-Linked Genes Inform Temporal Dynamics of Sex Chromosome Evolution and Suggest Candidate for Sex Determination. Mol. Biol. Evol. 38: 5345–5358. https://doi.org/10.1093/molbev/msab178
Ellegren H., 2011 Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat. Rev. Genet. 12: 157–166. https://doi.org/10.1038/nrg2948
Gammerdinger W. J., M. A. Toups, and B. Vicoso, 2020 Disagreement in F ST estimators: A case study from sex chromosomes. Mol. Ecol. Resour. 20: 1517–1525. https://doi.org/10.1111/1755-0998.13210
Gilbert D. G., 2019 Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? bioRxiv 829184; doi: https://doi.org/10.1101/829184.
Grabherr M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, *et al.*, 2011 Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29: 644–652. https://doi.org/10.1038/nbt.1883
Guan D., S. A. McCarthy, J. Wood, K. Howe, Y. Wang, *et al.*, 2020 Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36: 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025
Guindon S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, *et al.*, 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59: 307–321. https://doi.org/10.1093/sysbio/syq010
Haag C. R., L. Theodosiou, R. Zahab, and T. Lenormand, 2017 Low recombination rates in sexual species and sex–asex transitions. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160461. https://doi.org/10.1098/rstb.2016.0461
Han X., Y. Ren, X. Ouyang, B. Zhang, and L. Sui, 2021 Construction of a high-density genetic linkage map and QTL mapping for sex and growth traits in *Artemia franciscana*. Aquaculture 540: 736692. https://doi.org/10.1016/j.aquaculture.2021.736692
Handley L.-J. L., H. Ceplitis, and H. Ellegren, 2004 Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics 167: 367–376.
Huylmans A. K., A. Macon, and B. Vicoso, 2017 Global Dosage Compensation Is Ubiquitous in Lepidoptera, but Counteracted by the Masculinization of the Z Chromosome. Mol. Biol. Evol. 34: 2637–2649. https://doi.org/10.1093/molbev/msx190
Huylmans A. K., M. A. Toups, A. Macon, W. J. Gammerdinger, and B. Vicoso, 2019 Sex-Biased Gene Expression and Dosage Compensation on the *Artemia franciscana* Z-Chromosome. Genome Biol. Evol. 11: 1033–1044. https://doi.org/10.1093/gbe/evz053
Huylmans A. K., A. Macon, F. Hontoria, and B. Vicoso, 2021 Transitions to asexuality and evolution of gene expression in *Artemia* brine shrimp. Proc. Biol. Sci. 288: 20211720. https://doi.org/10.1098/rspb.2021.1720
Jaquiéry J., S. Stoeckel, C. Larose, P. Nouhaud, C. Rispe, *et al.*, 2014 Genetic Control of Contagious Asexuality in the Pea Aphid. PLOS Genet. 10: e1004838. https://doi.org/10.1371/journal.pgen.1004838

Jo E., S. J. Lee, E. Choi, J. Kim, S. G. Lee, *et al.*, 2021a Whole genome survey and microsatellite motif identification of *Artemia franciscana*. Biosci. Rep. 41: BSR20203868. https://doi.org/10.1042/BSR20203868
Jo E., S.-J. Lee, E. Choi, J. Kim, J.-H. Lee, *et al.*, 2021b Sex-Biased Gene Expression and Isoform Profile of Brine Shrimp *Artemia franciscana* by Transcriptome Analysis. Anim. Open Access J. MDPI 11: 2630. https://doi.org/10.3390/ani11092630
Kampfraath A. A., T. P. Dudink, K. Kraaijeveld, J. Ellers, and Z. V. Zizzari, 2020 Male Sexual Trait Decay in Two Asexual Springtail Populations Follows Neutral Mutation Accumulation Theory. Evol. Biol. 47: 285–292. https://doi.org/10.1007/s11692-020-09511-z
Katoh K., K. Misawa, K. Kuma, and T. Miyata, 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30: 3059–3066. https://doi.org/10.1093/nar/gkf436
Katsuma S., K. Shoji, Y. Sugano, Y. Suzuki, and T. Kiuchi, 2019 Masc-induced dosage compensation in silkworm cultured cells. FEBS Open Bio 9: 1573–1579. https://doi.org/10.1002/2211-5463.12698
Kent W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res. 12: 656–664. https://doi.org/10.1101/gr.229202
Kiuchi T., H. Koga, M. Kawamoto, K. Shoji, H. Sakai, *et al.*, 2014 A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature 509: 633–636. https://doi.org/10.1038/nature13315
Kofler R., R. V. Pandey, and C. Schlötterer, 2011 PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27: 3435–3436. https://doi.org/10.1093/bioinformatics/btr589
Kolmogorov M., J. Yuan, Y. Lin, and P. A. Pevzner, 2019 Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37: 540–546. https://doi.org/10.1038/s41587-019-0072-8
Lahn B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science 286: 964–967. https://doi.org/10.1126/science.286.5441.964
Langmead B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359. https://doi.org/10.1038/nmeth.1923
Lenormand T., F. Fyon, E. Sun, and D. Roze, 2020 Sex Chromosome Degeneration by Regulatory Evolution. Curr. Biol. 30: 3001-3006.e5. https://doi.org/10.1016/j.cub.2020.05.052
Lenormand T., and D. Roze, 2022 Y recombination arrest and degeneration in the absence of sexual dimorphism. Science 375: 663–666. https://doi.org/10.1126/science.abj1813
Letunic I., and P. Bork, 2019 Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47: W256–W259. https://doi.org/10.1093/nar/gkz239
Li H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Li H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509
Li D., C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam, 2015 MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31: 1674–1676. https://doi.org/10.1093/bioinformatics/btv033
Li H., 2016 Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32: 2103–2110. https://doi.org/10.1093/bioinformatics/btw152
Luo R., B. Liu, Y. Xie, Z. Li, W. Huang, *et al.*, 2012 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1: 2047-217X-1–18. https://doi.org/10.1186/2047-217X-1-18
Lynch M., A. Seyfert, B. Eads, and E. Williams, 2008 Localization of the Genetic Determinants of Meiosis Suppression in *Daphnia pulex*. Genetics 180: 317–327.

https://doi.org/10.1534/genetics.107.084657
Maccari M., A. Gómez, F. Hontoria, and F. Amat, 2013a Functional rare males in diploid parthenogenetic *Artemia*. J. Evol. Biol. 26: 1934–1948. https://doi.org/10.1111/jeb.12191
Maccari M., F. Amat, and A. Gómez, 2013b Origin and Genetic Diversity of Diploid Parthenogenetic *Artemia* in Eurasia. PLOS ONE 8: e83348. https://doi.org/10.1371/journal.pone.0083348
Maccari M., F. Amat, F. Hontoria, and A. Gómez, 2014 Laboratory generation of new parthenogenetic lineages supports contagious parthenogenesis in *Artemia*. PeerJ 2: e439. https://doi.org/10.7717/peerj.439
Mank J. E., 2013 Sex chromosome dosage compensation: definitely not for everyone. Trends Genet. 29: 677–683. https://doi.org/10.1016/j.tig.2013.07.005
Manni M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov, 2021 BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. 38: 4647–4654. https://doi.org/10.1093/molbev/msab199
Meiklejohn C. D., E. L. Landeen, J. M. Cook, S. B. Kingan, and D. C. Presgraves, 2011 Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. PLoS Biol. 9: e1001126. https://doi.org/10.1371/journal.pbio.1001126
Mignerot L., K. Avia, R. Luthringer, A. P. Lipinska, A. F. Peters, *et al.*, 2019 A key role for sex chromosomes in the regulation of parthenogenesis in the brown alga Ectocarpus. PLoS Genet. 15: e1008211. https://doi.org/10.1371/journal.pgen.1008211
Mullon C., A. E. Wright, M. Reuter, A. Pomiankowski, and J. E. Mank, 2015 Evolution of dosage compensation under sexual selection differs between X and Z chromosomes. Nat. Commun. 6: 7720. https://doi.org/10.1038/ncomms8720
Neiman M., T. F. Sharbel, and T. Schwander, 2014 Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. J. Evol. Biol. 27: 1346–1359. https://doi.org/10.1111/jeb.12357
Nougué O., N. O. Rode, R. Jabbour-Zahab, A. Ségard, L.-M. Chevin, *et al.*, 2015 Automixis in *Artemia*: solving a century-old controversy. J. Evol. Biol. 28: 2337–2348. https://doi.org/10.1111/jeb.12757
Palmer D. H., T. F. Rogers, R. Dean, and A. E. Wright, 2019 How to identify sex chromosomes and their turnover. Mol. Ecol. 28: 4709–4724. https://doi.org/10.1111/mec.15245
Parraguez M., G. Gajardo, and J. A. Beardmore, 2009 The New World *Artemia* species A. franciscana and A. persimilis are highly differentiated for chromosome size and heterochromatin content. Hereditas 146: 93–103. https://doi.org/10.1111/j.1601-5223.2009.02109.x
Pennell M. W., J. E. Mank, and C. L. Peichel, 2018 Transitions in sex determination and sex chromosomes across vertebrate species. Mol. Ecol. 27: 3950–3963. https://doi.org/10.1111/mec.14540
Picard M. A. L., C. Cosseau, S. Ferré, T. Quack, C. G. Grevelding, *et al.*, 2018 Evolution of gene dosage on the Z-chromosome of schistosome parasites. eLife 7: e35684. https://doi.org/10.7554/eLife.35684
Picard M. A. L., B. Vicoso, S. Bertrand, and H. Escriva, 2021 Diversity of Modes of Reproduction and Sex Determination Systems in Invertebrates, and the Putative Contribution of Genetic Conflict. Genes 12: 1136. https://doi.org/10.3390/genes12081136
Ponnikas S., H. Sigeman, J. K. Abbott, and B. Hansson, 2018 Why Do Sex Chromosomes Stop Recombining? Trends Genet. 34: 492–503. https://doi.org/10.1016/j.tig.2018.04.001
Robinson J. T., D. Turner, N. C. Durand, H. Thorvaldsdóttir, J. P. Mesirov, *et al.*, 2018 Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst. 6: 256-258.e1. https://doi.org/10.1016/j.cels.2018.01.001
Rode N. O., R. Jabbour-Zahab, L. Boyer, É. Flaven, F. Hontoria, *et al.*, 2022 The origin of asexual brine shrimps. Am. Nat. https://doi.org/10.1086/720268
Rosin L. F., D. Chen, Y. Chen, and E. P. Lei, 2022 Dosage compensation in Bombyx mori is achieved by partial repression of both Z chromosomes in males. Proc. Natl. Acad. Sci. 119:

e2113374119. https://doi.org/10.1073/pnas.2113374119
Rovatsos M., and L. Kratochvíl, 2021 Evolution of dosage compensation does not depend on genomic background. Mol. Ecol. 30: 1836–1845. https://doi.org/10.1111/mec.15853
Ruan J., and H. Li, 2020 Fast and accurate long-read assembly with wtdbg2. Nat. Methods 17: 155–158. https://doi.org/10.1038/s41592-019-0669-3
Sainz-Escudero L., E. K. López-Estrada, P. C. Rodríguez-Flores, and M. García-París, 2021 Settling taxonomic and nomenclatural problems in brine shrimps, *Artemia* (Crustacea: Branchiopoda: Anostraca), by integrating mitogenomics, marker discordances and nomenclature rules. PeerJ 9: e10865. https://doi.org/10.7717/peerj.10865
Sandrock C., and C. Vorburger, 2011 Single-Locus Recessive Inheritance of Asexual Reproduction in a Parasitoid Wasp. Curr. Biol. 21: 433–437. https://doi.org/10.1016/j.cub.2011.01.070
Schwander T., and B. J. Crespi, 2009 Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in Timema stick insects. Evol. Int. J. Org. Evol. 63: 84–103. https://doi.org/10.1111/j.1558-5646.2008.00524.x
Schwentner M., S. Richter, D. C. Rogers, and G. Giribet, 2018 Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. Proc. Biol. Sci. 285: 20181524. https://doi.org/10.1098/rspb.2018.1524
Sigeman H., M. Strandh, E. Proux-Wéra, V. E. Kutschera, S. Ponnikas, *et al.*, 2021 Avian Neo-Sex Chromosomes Reveal Dynamics of Recombination Suppression and W Degeneration. Mol. Biol. Evol. 38: 5275–5291. https://doi.org/10.1093/molbev/msab277
Straub T., and P. B. Becker, 2007 Dosage compensation: the beginning and end of generalization. Nat. Rev. Genet. 8: 47–57. https://doi.org/10.1038/nrg2013
Talavera G., and J. Castresana, 2007 Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Syst. Biol. 56: 564–577. https://doi.org/10.1080/10635150701472164
Toman J., and J. Flegr, 2018 General environmental heterogeneity as the explanation of sexuality? Comparative study shows that ancient asexual taxa are associated with both biotically and abiotically homogeneous environments. Ecol. Evol. 8: 973–991. https://doi.org/10.1002/ece3.3716
Törönen P., A. Medlar, and L. Holm, 2018 PANNZER2: a rapid functional annotation web server. Nucleic Acids Res. 46: W84–W88. https://doi.org/10.1093/nar/gky350
Turner S. D., 2018 qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J. Open Source Softw. 3: 731. https://doi.org/10.21105/joss.00731
Van Stappen G., 2002 Zoogeography, pp. 171–224 in *Artemia: Basic and Applied Biology*, Biology of Aquatic Organisms. edited by Abatzopoulos Th. J., Beardmore J. A., Clegg J. S., Sorgeloos P. Springer Netherlands, Dordrecht.
Vaser R., I. Sović, N. Nagarajan, and M. Šikić, 2017 Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27: 737–746. https://doi.org/10.1101/gr.214270.116
Vibranovski M. D., H. F. Lopes, T. L. Karr, and M. Long, 2009 Stage-Specific Expression Profiling of *Drosophila* Spermatogenesis Suggests that Meiotic Sex Chromosome Inactivation Drives Genomic Relocation of Testis-Expressed Genes. PLOS Genet. 5: e1000731. https://doi.org/10.1371/journal.pgen.1000731
Vicoso B., and D. Bachtrog, 2009 Progress and prospects toward our understanding of the evolution of dosage compensation. Chromosome Res. 17: 585. https://doi.org/10.1007/s10577-009-9053-y
Vicoso B., J. J. Emerson, Y. Zektser, S. Mahajan, and D. Bachtrog, 2013 Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11: e1001643. https://doi.org/10.1371/journal.pbio.1001643
Vicoso B., 2019 Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat. Ecol. Evol. 3: 1632–1641. https://doi.org/10.1038/s41559-019-1050-8
Walters J. R., and T. J. Hardcastle, 2011 Getting a full dose? Reconsidering sex chromosome

dosage compensation in the silkworm, Bombyx mori. Genome Biol. Evol. 3: 491–504. https://doi.org/10.1093/gbe/evr036
Wang M., and L. Kong, 2019 pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinformatics 20: 28. https://doi.org/10.1186/s12859-019-2597-8
Willforss J., A. Chawade, and F. Levander, 2019 NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. J. Proteome Res. 18: 732–740. https://doi.org/10.1021/acs.jproteome.8b00523
Wright A. E., R. Dean, F. Zimmer, and J. E. Mank, 2016 How to make a sex chromosome. Nat. Commun. 7: 12087. https://doi.org/10.1038/ncomms12087
Yagound B., K. A. Dogantzis, A. Zayed, J. Lim, P. Broekhuyse, *et al.*, 2020 A Single Gene Causes Thelytokous Parthenogenesis, the Defining Feature of the Cape Honeybee Apis mellifera capensis. Curr. Biol. 30: 2248-2259.e6. https://doi.org/10.1016/j.cub.2020.04.033
Zeileis A., and G. Grothendieck, 2005 zoo: S3 Infrastructure for Regular and Irregular Time Series. J. Stat. Softw. 14: 1–27. https://doi.org/10.18637/jss.v014.i06
Zhou Q., J. Zhang, D. Bachtrog, N. An, Q. Huang, *et al.*, 2014 Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346: 1246338. https://doi.org/10.1126/science.1246338
Zhou C., S. A. McCarthy, and R. Durbin, 2022 YaHS: yet another Hi-C scaffolding tool. bioRxiv 2022.06.09.495093; doi: https://doi.org/10.1101/2022.06.09.495093

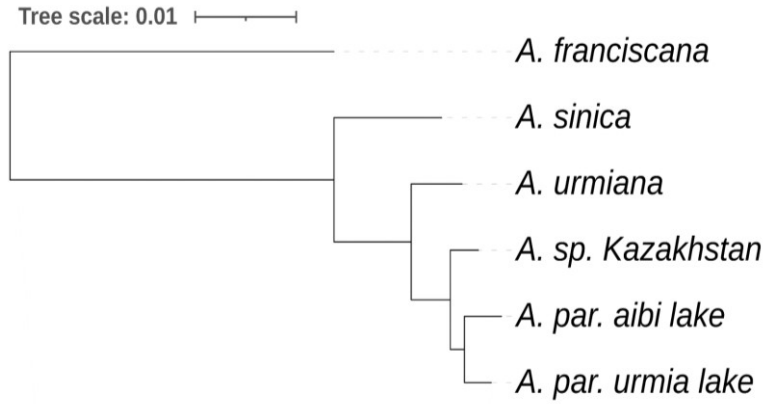
Supplementary Material

Supplementary Datasets

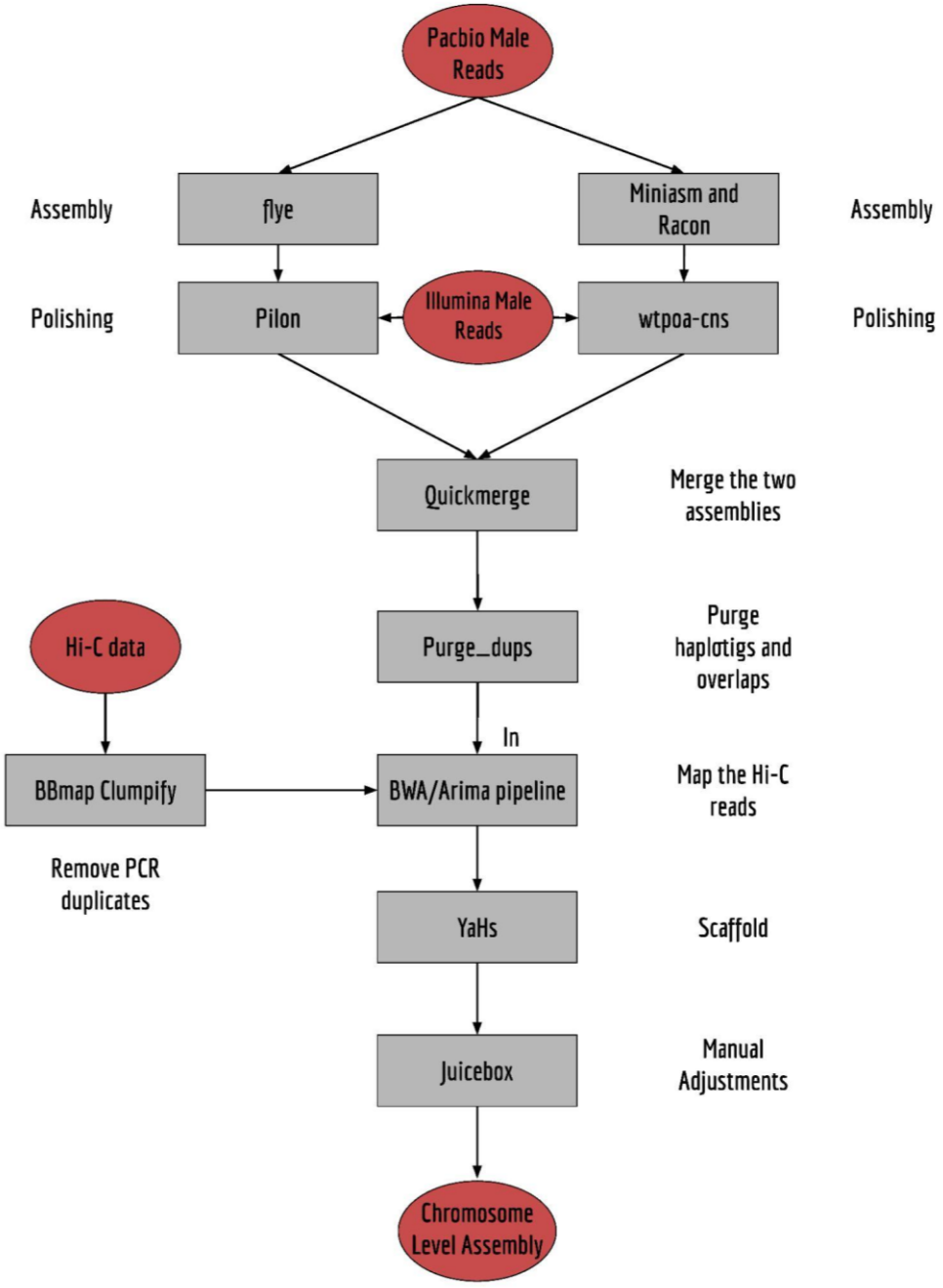
Sup. Dataset 1: Genome assemblies: *A. sinica* male high quality assembly, *A. sp. Kazakhstan* male draft assembly
Sup. Dataset 2: Male transcriptome assemblies for *A. sinica* and *A. franciscana*
Sup. Dataset 3: Male and female coverage for *A. sinica*, *A. sp. Kazakhstan*, *A. urmiana*, and *A. parthenogenetica* females and rare male.
Sup. Dataset 4: *Artemia sinica* Male:female F_{ST} per 1Kb window
Sup. Dataset 5: FASTA file with candidate W scaffolds
Sup. Dataset 6: Candidate W-derived transcripts and alignments
Sup. Dataset 7: Gene expression with genomic location
Sup. Dataset 8: VCF for asexual female and rare male
Sup. Dataset 9: F_{ST} between backcrossed asexual and control females (pooled analysis
Sup. Dataset 10: VCF of backcrossed asexual and control females (individual analysis using *A. sp. Kazakhstan* as the reference), and inferred ancestry
Sup. Dataset 11: GO and DE annotations of all the *Artemia sinica* transcripts and their locations in the *Artemia sinica* male genome.

The supplementary datasets are available here: https://doi.org/10.15479/AT:ISTA:11653.

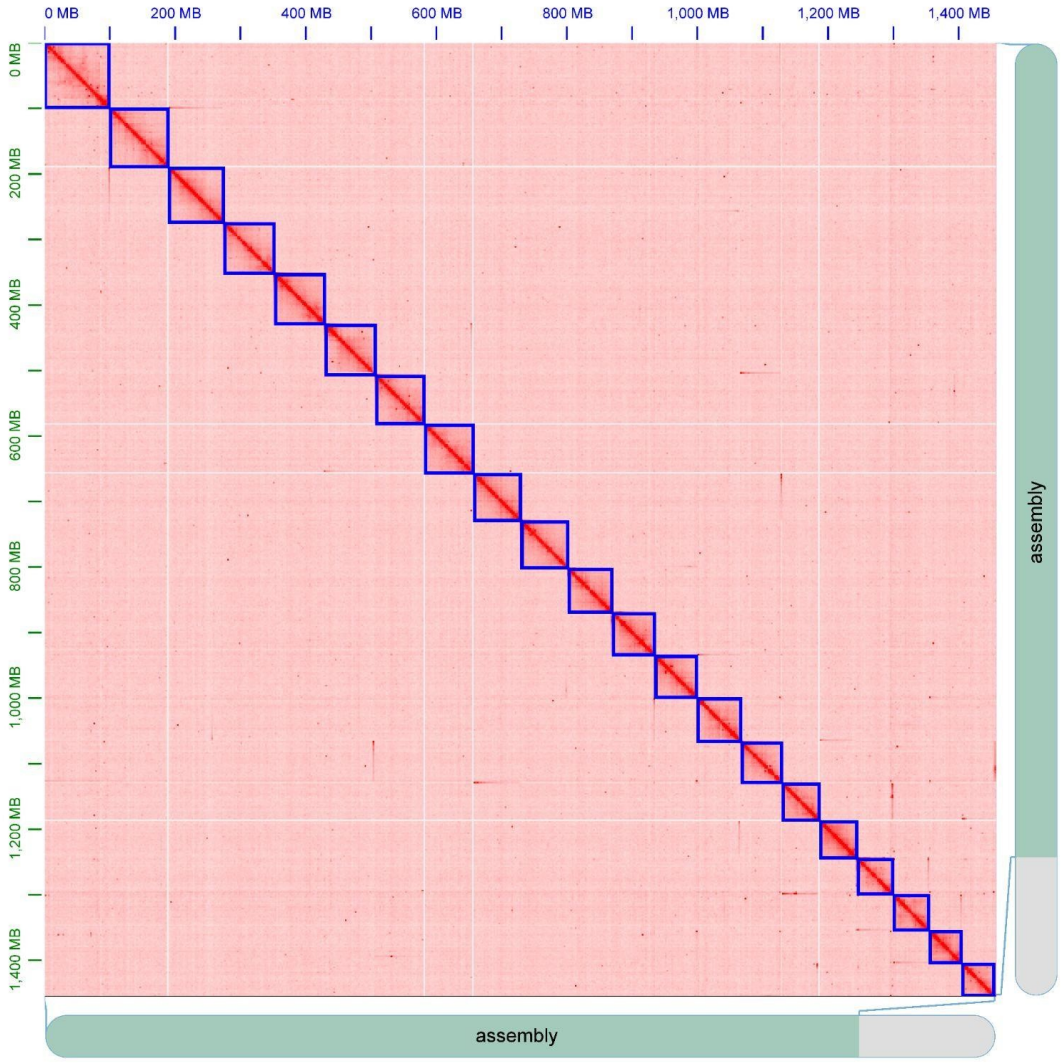
Supplementary Figures



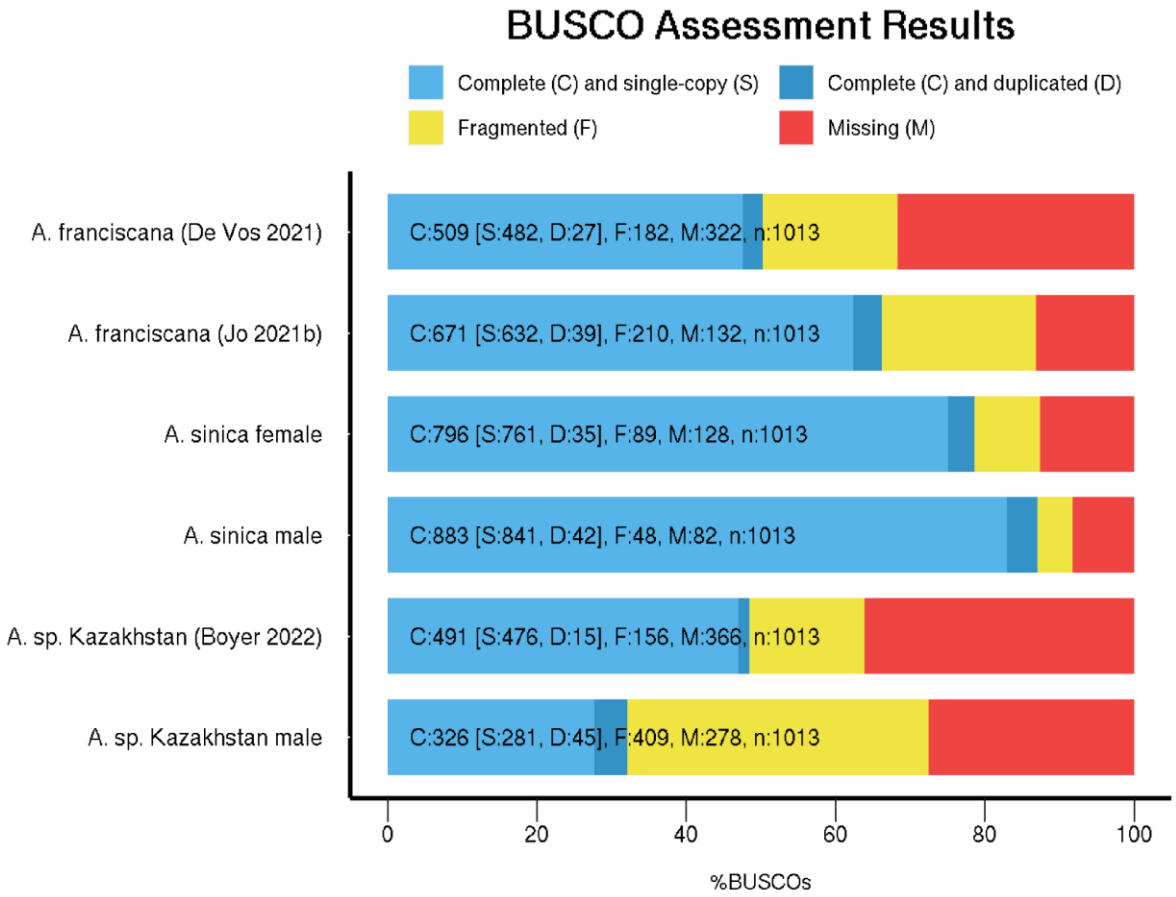
Supplementary Figure 1. Phylogeny of the lineages used in the study. The phylogeny was adapted from (Huylmans et al 2021).



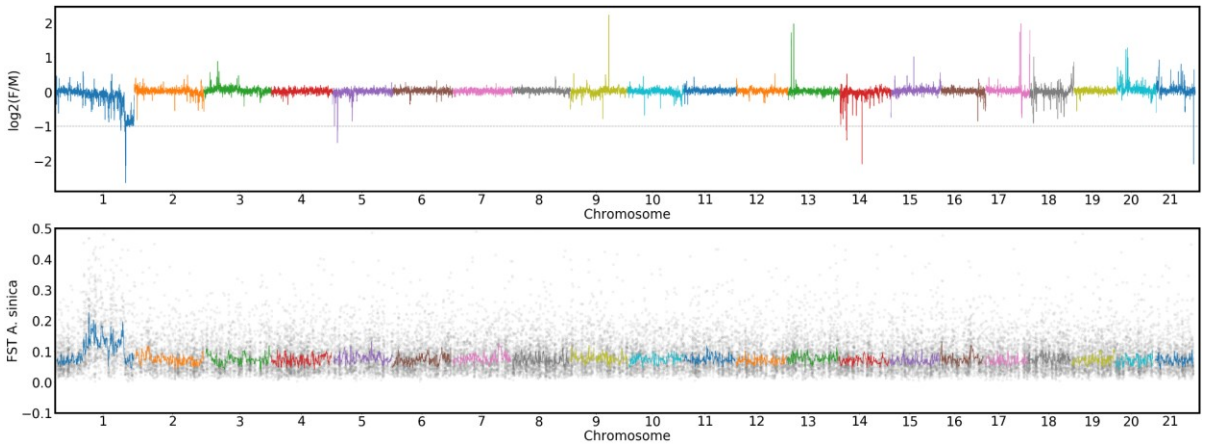
Supplementary Figure 2. Pipeline of *A. sinica* male genome assembly.



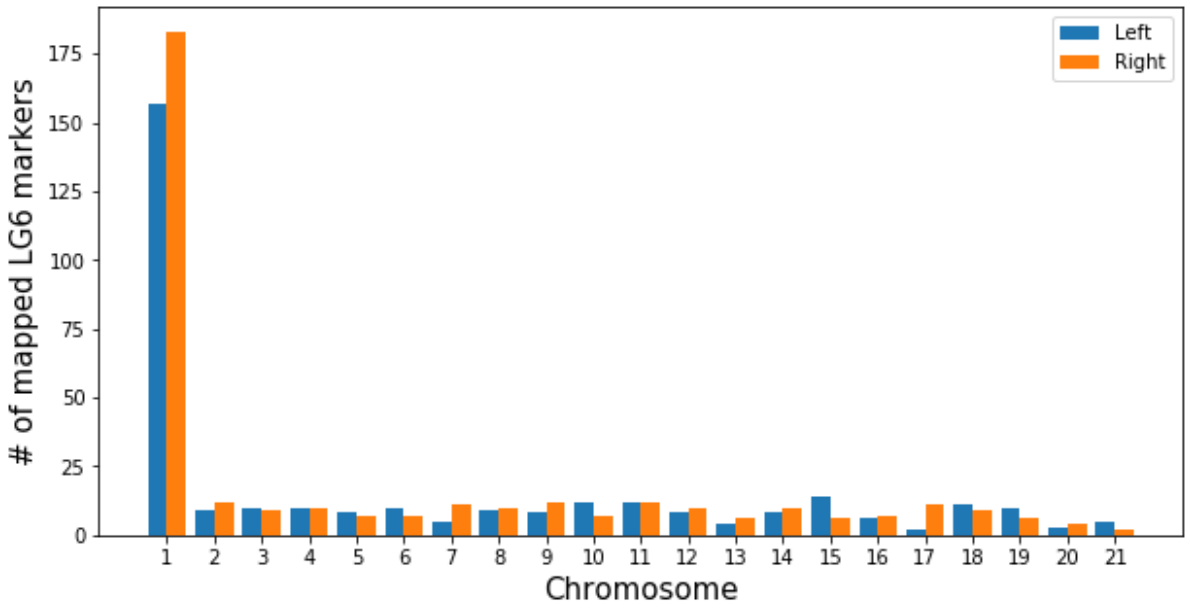
Supplementary Figure 3. Heatmap of the Hi-C contact matrix on the *A. sinica* genome assembly. Blue boxes represent the largest 21 scaffolds, which correspond to the 21 chromosomes.



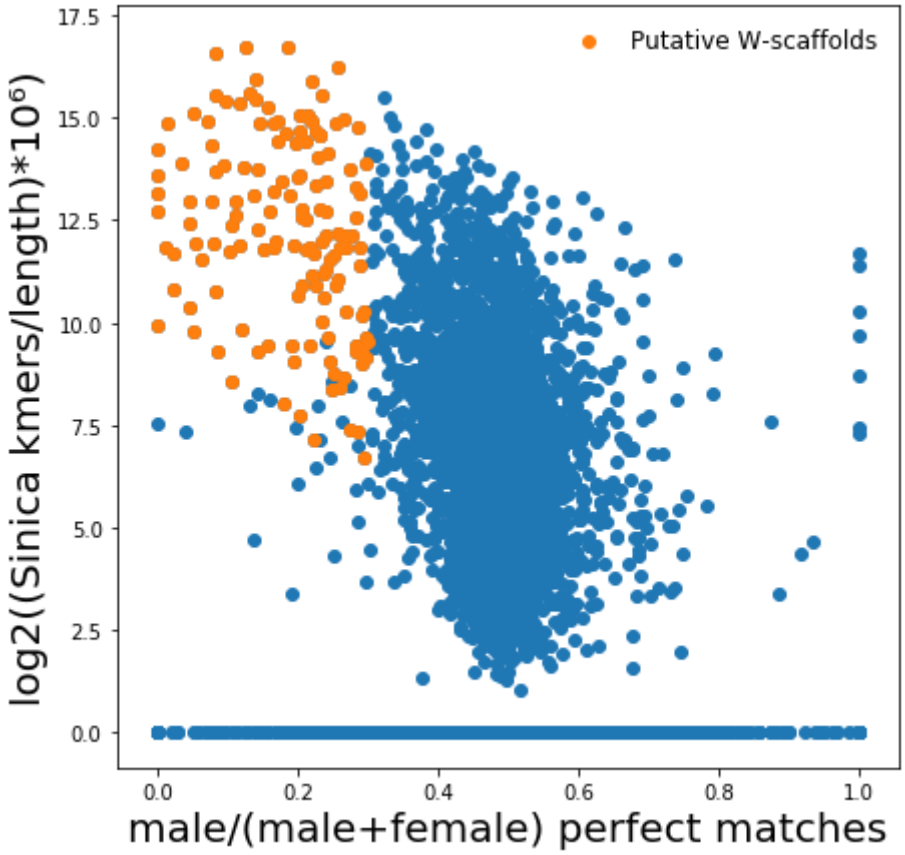
Supplementary Figure 4. Comparison of BUSCO scores of available *Artemia* genome assemblies.



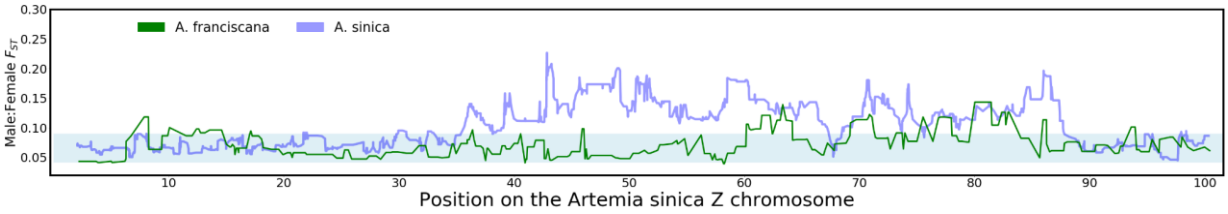
Supplementary Figure 5. Coverage patterns (log2(female/male)) and male-female F_{ST} for all the *Artemia sinica* chromosomes.



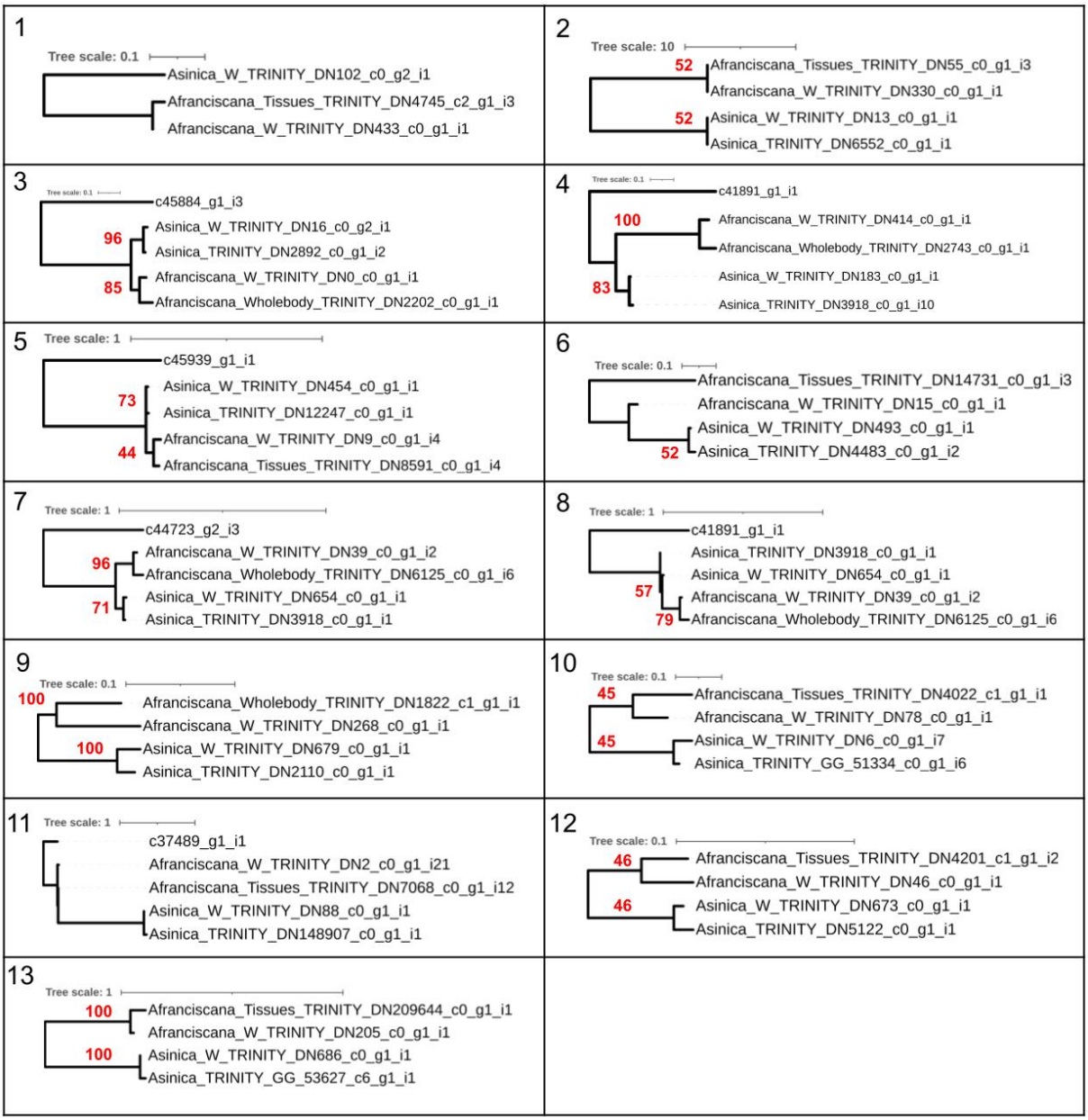
Supplementary Figure 6. Mapping of *A. franciscana* LG6 markers to the *A. sinica* chromosomes. The two paired-ends of the markers were mapped individually ("Left" and "Right" markers in the plot).



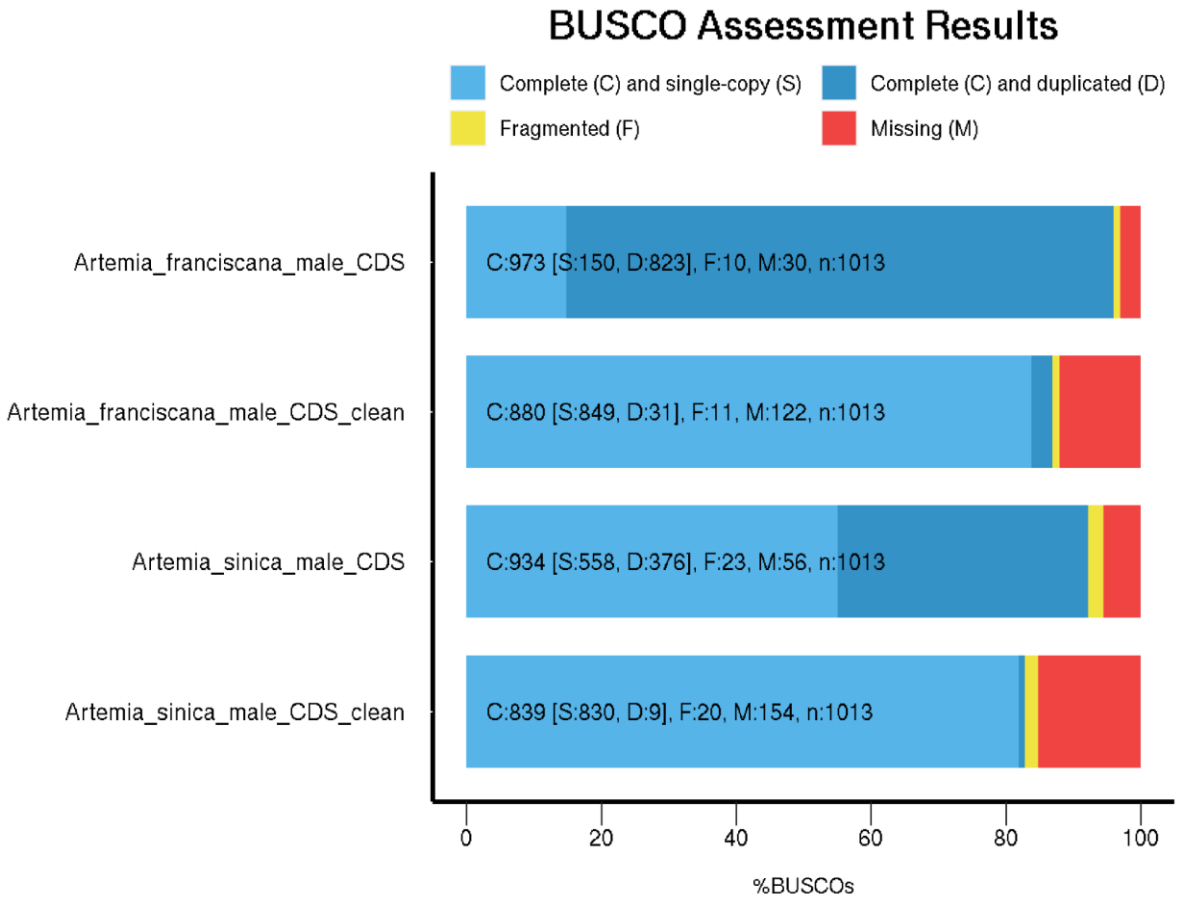
Supplementary Figure 7. Putative W-derived scaffolds in the *A. sinica* female genome assembly. Scaffolds harboring more than 5 female-specific Kmers, and with more perfectly matching genomic reads in two female samples than in two male samples (male/(male+female) <=0.3), were classified as putatively W-derived (orange dots on the plot).



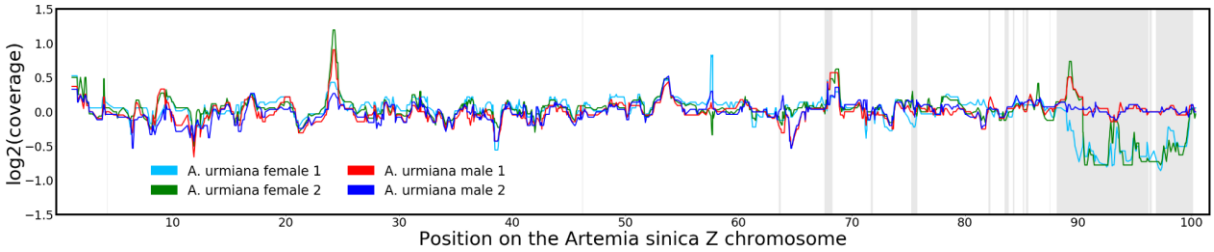
Supplementary Figure 8. *A. franciscana* male:female F_{ST}



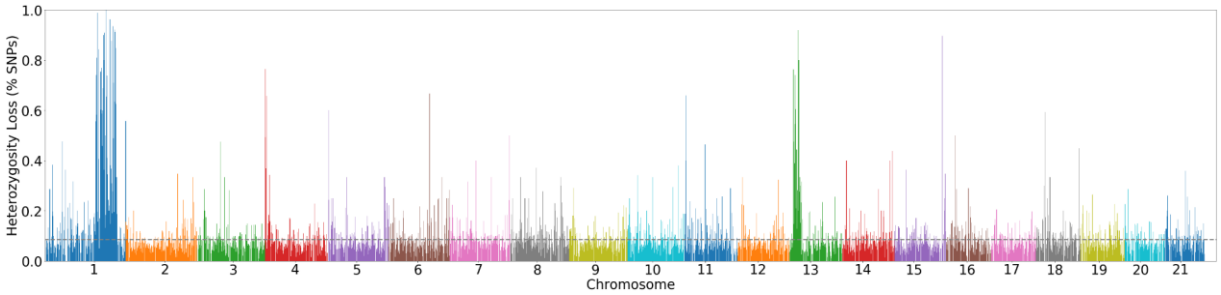
Supplementary Figure 9. Phylogenetic trees of candidate W-genes shared by both *A. franciscana* and *A. sinica*. The bootstrap values are in red (based on 100 bootstraps).



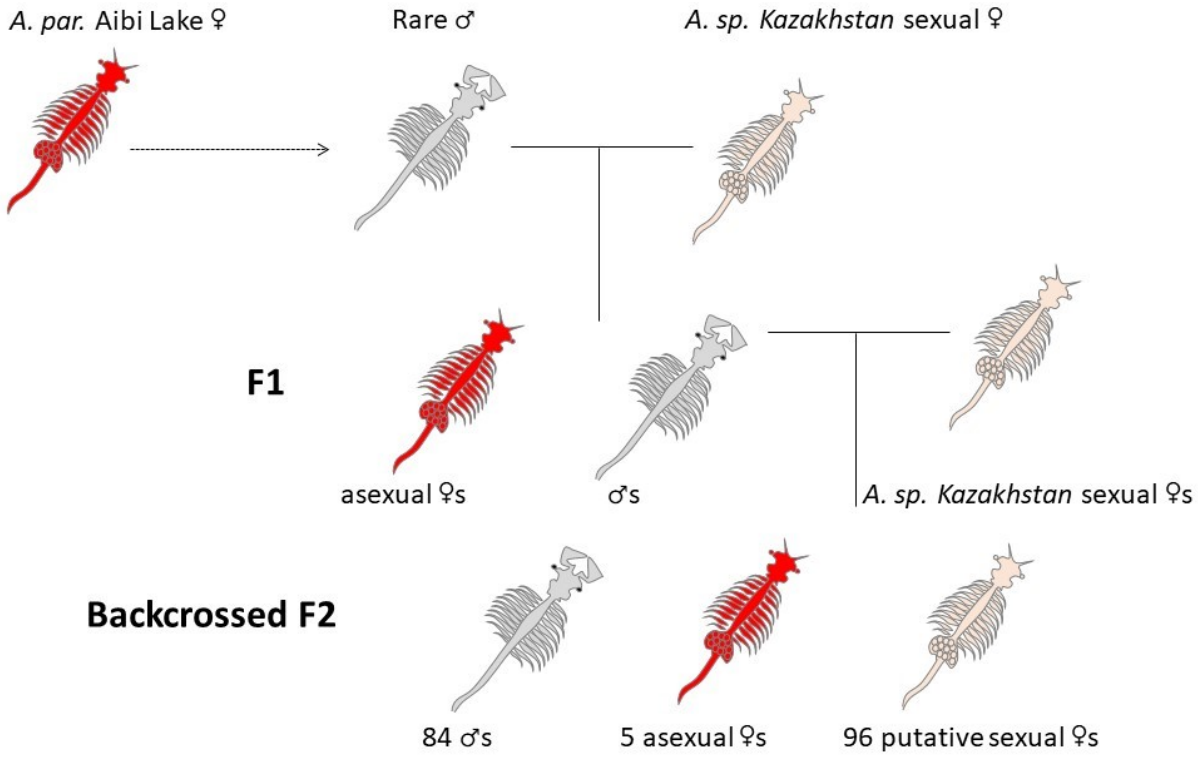
Supplementary Figure 10. BUSCO assessment of new transcriptome assemblies.



Supplementary Figure 11. Coverage patterns in *A. urmiana*.

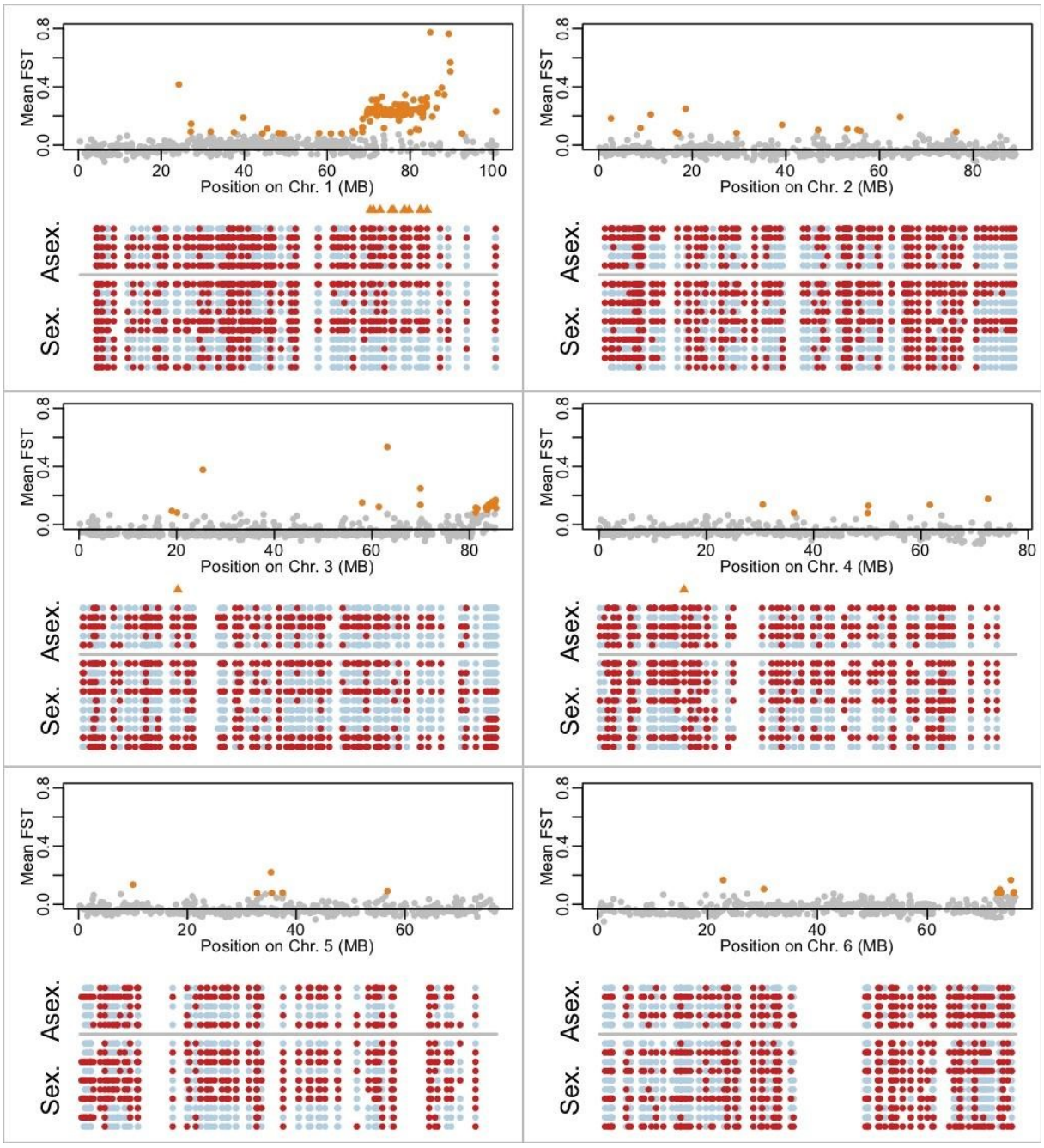


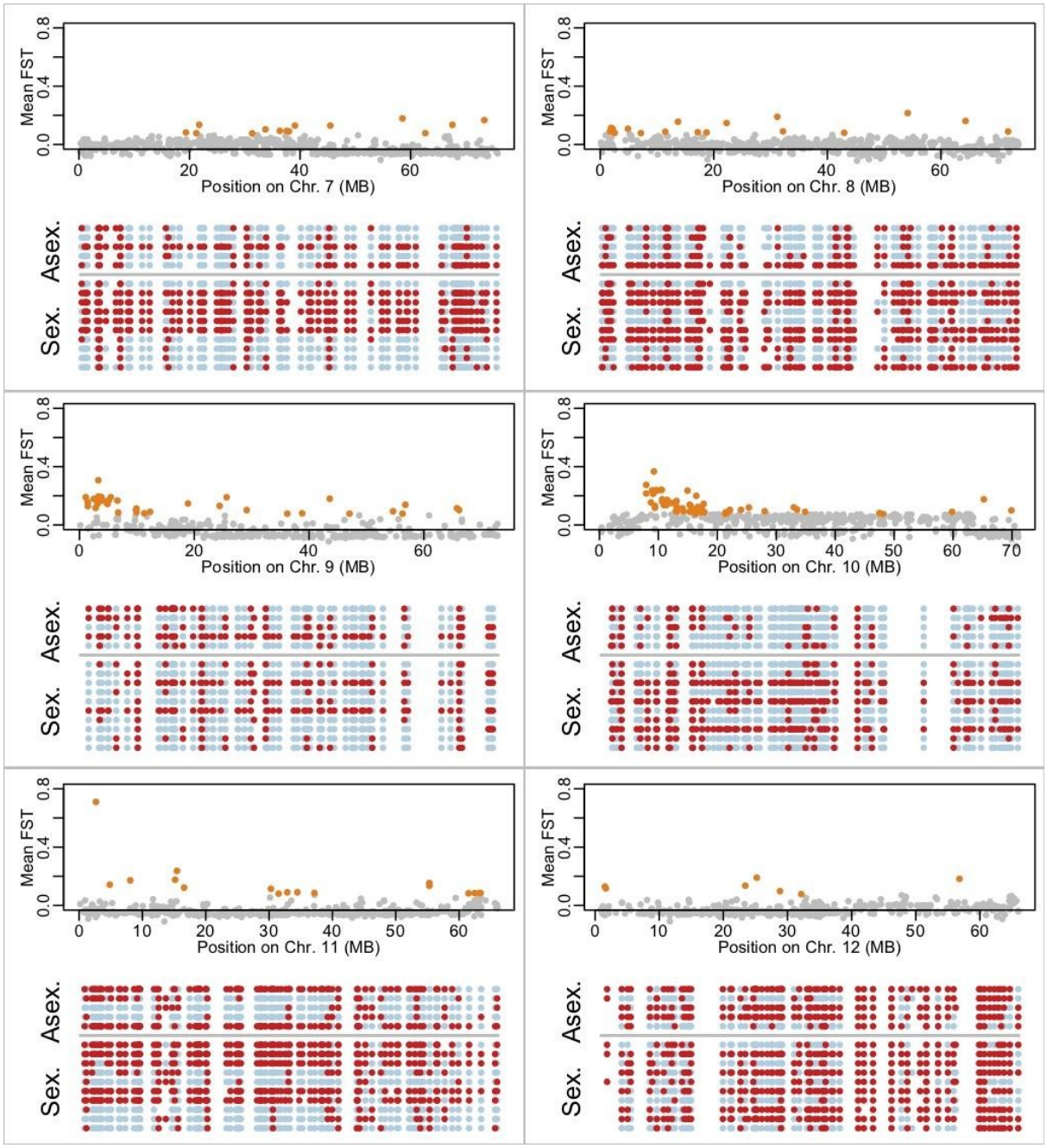
Supplementary Figure 12. Loss of heterozygosity across the genome in the rare male (compared to its *A. parthenogenetica* sister).

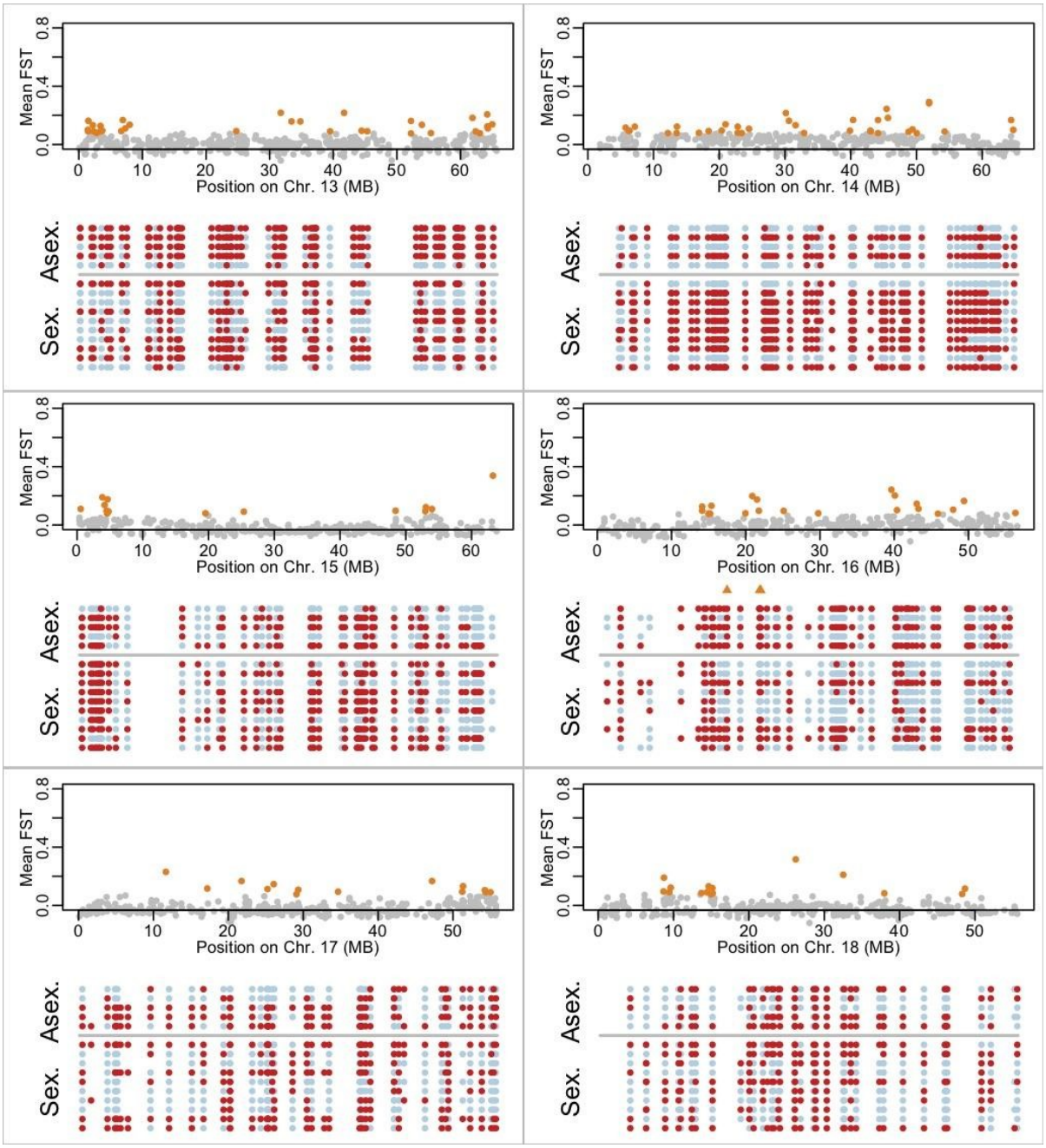


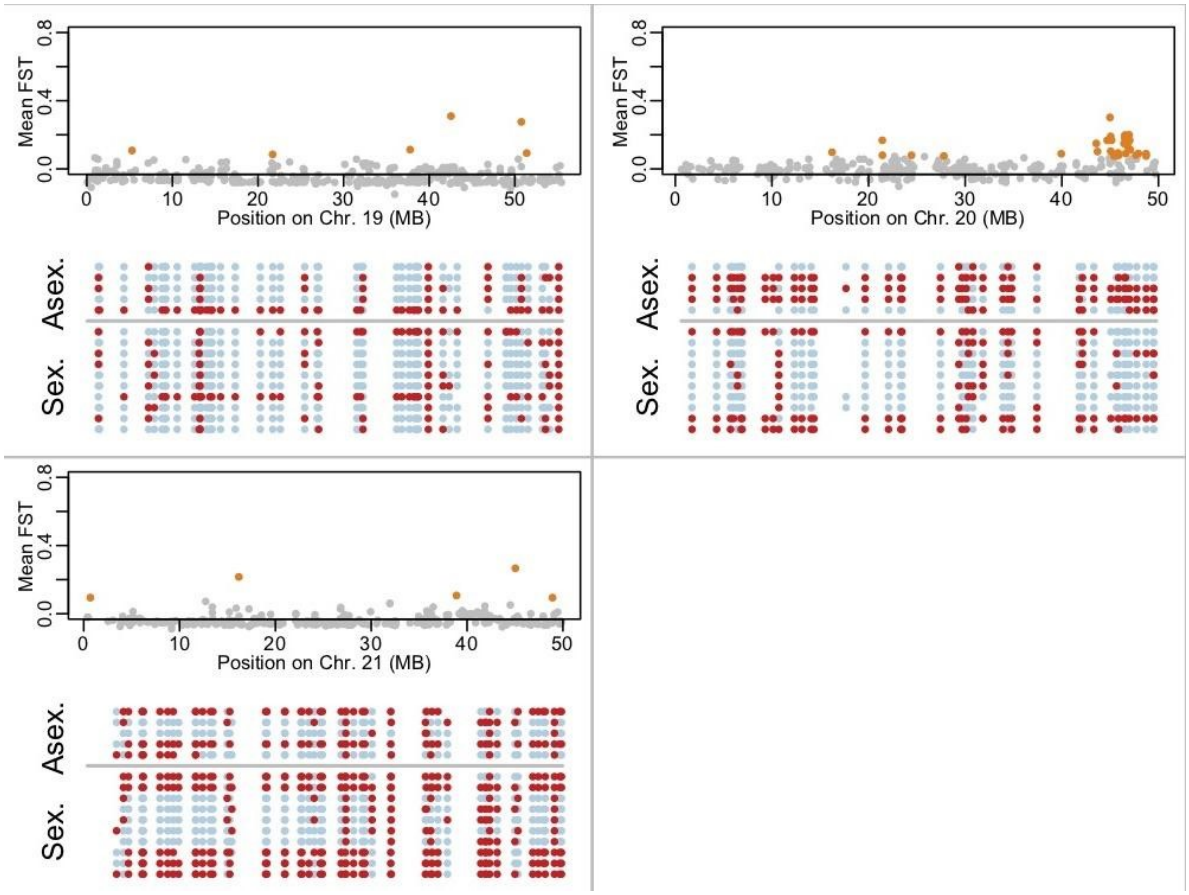
Supplementary Figure 13. Crossing scheme for identifying genomic regions linked to the spread of asexuality.

Supplementary Figure 14 (next 4 pages). Asexual:sexual genetic differentiation and *A. parthenogenetic Aibi Lake* ancestry when using individual genotypes called with the *A. sp. Kazakhstan* genome as the reference. For each chromosome, the top panel shows the mean F_{ST} between the 5 F2 asexual females and the 10 control putative sexual females for windows of 10KB (orange dots denote windows above the 95% percentile). The bottom panel shows the inferred ancestry of each scaffold (blue denotes *A. sp. Kazakhstan* ancestry only, red denotes both *A. sp. Kazakhstan* and *A. parthenogenetica Aibi Lake* ancestry), for each of the 5 F2 asexual females ("Asex." in the figure) and the 10 control putative sexual females ("Sex." in the figure). Orange triangles denote regions consistent with an asexuality locus, i.e. scaffolds for which all asexuals but not all putative sexuals have *Aibi Lake* ancestry. All plots were made using only SNPs that are fixed between *A. sp. Kazakhstan* and *A. parthenogenetica Aibi Lake*.




Mean FST
0.0
0.4
0.8
0
20
40
60
Position on Chr. 7 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
20
40
60
Position on Chr. 8 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
20
40
60
Position on Chr. 9 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
10
20
30
40
50
60
70
Position on Chr. 10 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
10
20
30
40
50
60
Position on Chr. 11 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
10
20
30
40
50
60
Position on Chr. 12 (MB)
Asex.
Sex.


Mean FST
0.0
0.4
0.8
0 10 20 30 40 50 60
Position on Chr. 13 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0 10 20 30 40 50 60
Position on Chr. 14 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0 10 20 30 40 50 60
Position on Chr. 15 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0 10 20 30 40 50
Position on Chr. 16 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0 10 20 30 40 50
Position on Chr. 17 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0 10 20 30 40 50
Position on Chr. 18 (MB)
Asex.
Sex.


Mean FST
0.0
0.4
0.8
0
10
20
30
40
50
Position on Chr. 19 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
10
20
30
40
50
Position on Chr. 20 (MB)
Asex.
Sex.
Mean FST
0.0
0.4
0.8
0
10
20
30
40
50
Position on Chr. 21 (MB)
Asex.
Sex.

Supplementary Tables

Supplementary Table 1: Statistics for the different assembly steps of the male *A. sinica* genome.

	Flye (+Pilon)	**Miniasm (Racon +wtpoa-cns)**	**Quickmerge**	**Purge-dups**	**Yahs + juicebox**
size	1809023064	1798524669	1870672883	1700852953	1701024053
n	13618	5310	2946	2027	1213
N50	791316	570290	1969269	2176199	67194031
N90	86138	141470	255797	408149	1655253
largest	9906702	4244182	17718575	17718575	100949155
Average	132840.58	338705.21	634987.40	839098.64	1402328.16
N_count	14543	0	9276	9176	180276
Gaps	160	0	104	103	1812

Supplementary Table 2: The coordinates of the different regions of the Z chromosome and the number of the putative W and Z transcripts in those regions.

Region	Start Coordinate (>=)	End Coordinate (<)	*Artemia sinica* W candidates	*Artemia sinica* Z	*Artemia franciscana* W candidates	*Artemia franciscana* Z
PAR	0	35322500	9	218	8	210
S1b	35322500	63525001	29	242	8	244
S1a	63525001	88115001	117	179	144	176
S0	88115001	100585001	16	84	7	91
unclassified	100585001	100949155	9	4	1	1

Supplementary Table 3: Bonferroni-corrected p-values from the Wilcoxon tests comparing the expression of the autosomes and the different regions of the Z chromosome.

	Gonads p-values	Heads p-values	Thoraces p-values
autosomes-PAR	0.90912	1	1
autosomes-S1b	0.07039	1	1
autosomes-S1a	1	0.13204	1
autosomes-S0	**0.00067**	0.19521	0.55028

Supplementary Table 4: Statistics for the two steps of the assembly of the *A. sp. Kazakhstan* Male Genome.

	Megahit	**SOAPdenovo-fusion**
size	1225371478	1229083815
n	910198	726632
N50	3237	4833
N90	440	503
largest	902753	996477
Average	1346.27	1691.48
N_count	0	10408426
Gaps	0	183782

Supplementary Table 5: Counts of asexual females, control females, and males, among the F2 progeny of the rare male / A. sp. Kazakhstan backcrosses.

Cross	total individuals	males	females	dead before sexing	Asexual females	Control females sequenced
1	22	10	7	5	1	2
2	66	21	35	10	1	2
3	17	5	5	7	1	2
4	38	18	18	2	0	0
5	41	20	16	5	0	0
6	45	10	20	15	2	4
total	229	84	101	44	5	10

Supplementary Table 6: Genomic samples generated for this study and what steps of the analysis they were used in. This table is available here.

Supplementary Table 7: Published RNA samples used in this study and what steps of the analysis they were used in. This table is available here.

Chapter 2.2: Chromosome-level assembly of *Artemia franciscana* sheds light on sex-chromosome differentiation

Authors:

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

1 Institute of Science and Technology Austria (ISTA), Klosterneuburg 3400, Austria
co-contributors
@ co-corresponding

Key words:
sex chromosome evolution, genome assembly, dosage compensation

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, are 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 *Artemia franciscana* (Kellogg 1906), from the Great Salt Lake, USA. The genome is 1GB, 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 F_{ST} 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 the 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 1200 species (Castellucci et al. 2022). They live in saline/hyper-saline 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 & Macrae 2002). Their adaptability, ease of rearing, and high nutritional value have made them very popular in the aquarium trade and

aquaculture industry (Lavens & 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 two scaffold-level assemblies (Jo et al. 2021; De Vos 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 & 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 two 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 *Artemia 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 & Clark 2013; Elkrewi et al. 2021), to remove CCS reads originating from the W chromosome (25,784 reads, 0.52%, Supplementary Figure 1). 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 Figure 2).

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 Figure 3), and anchored them using the 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 Figure 4). 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 Figure 5), resulting in 506 putative W sequences, which were added to the assembly for the downstream analysis.

The final assembly has 2118 scaffolds and an N50 of 43 MB, with 81% of the assembly in the 21 linkage groups (Supplementary Table 1). We ran BUSCO to assess the completeness of the genome, and 88.5% of BUSCOs were assembled completely, with less than 7% missing (Figure 1 A). We also checked for contamination using blobtools, and the results show that there is less than 1% bacterial contamination and most sequences map to Arthropoda (Figure 1 B).

As another quality check, we compared our assembly to the *Artemia sinica* chromosomes. Both the *A. franciscana* and *A. sinica* genomes were annotated using RNA and protein evidence (as described in the Genome Annotation section). The annotations were used to examine and visualize the synteny between the two genomes using GENESPACE (Lovell et al. 2022). As Figure 1 C shows, the genomes are highly syntenic, with no evidence for any large-scale rearrangements.

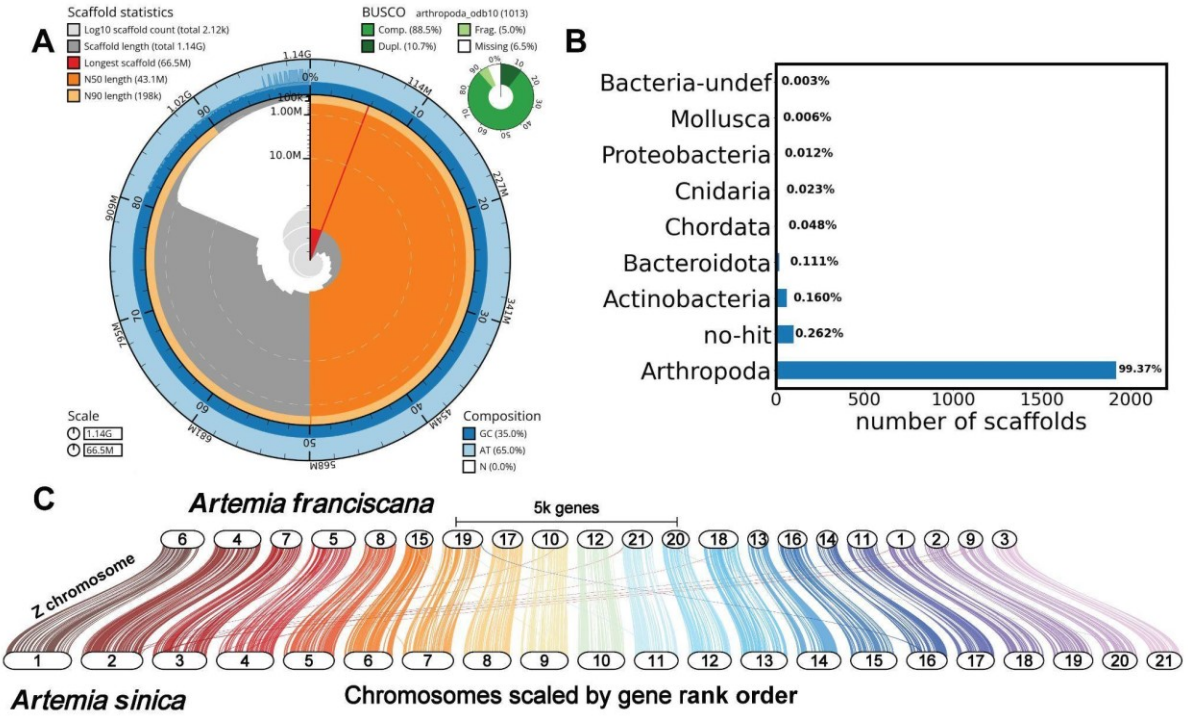


Figure 1: Genome assembly and synteny. A) Genome assembly statistics, GC content, and BUSCO score using the arthropoda_odb10 dataset. B) Barplot showing the number of sequences that map to the different phyla in the nt blast database and the percentage of the total assembly length they represent. C) Synteny between the *A. franciscana* and *A. sinica* genomes.

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 non-recombining but undifferentiated region (Huylmans et al. 2019). We estimated the coverage across the genome using short-read male and female DNA in windows of 10000 bp (Supplementary Figure 6), and then used the ratio of female to male coverage to identify regions that have become well differentiated between the Z (linkage group 6) and W chromosomes (Figure 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. However, 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 Figure 7). 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 two sides of the differentiated region (Figure 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 three 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 two 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 two chromosomes. The ZW pairs in *A. franciscana* S0 map to the S1a stratum of *A. sinica,* suggesting that they might not be ancestral.

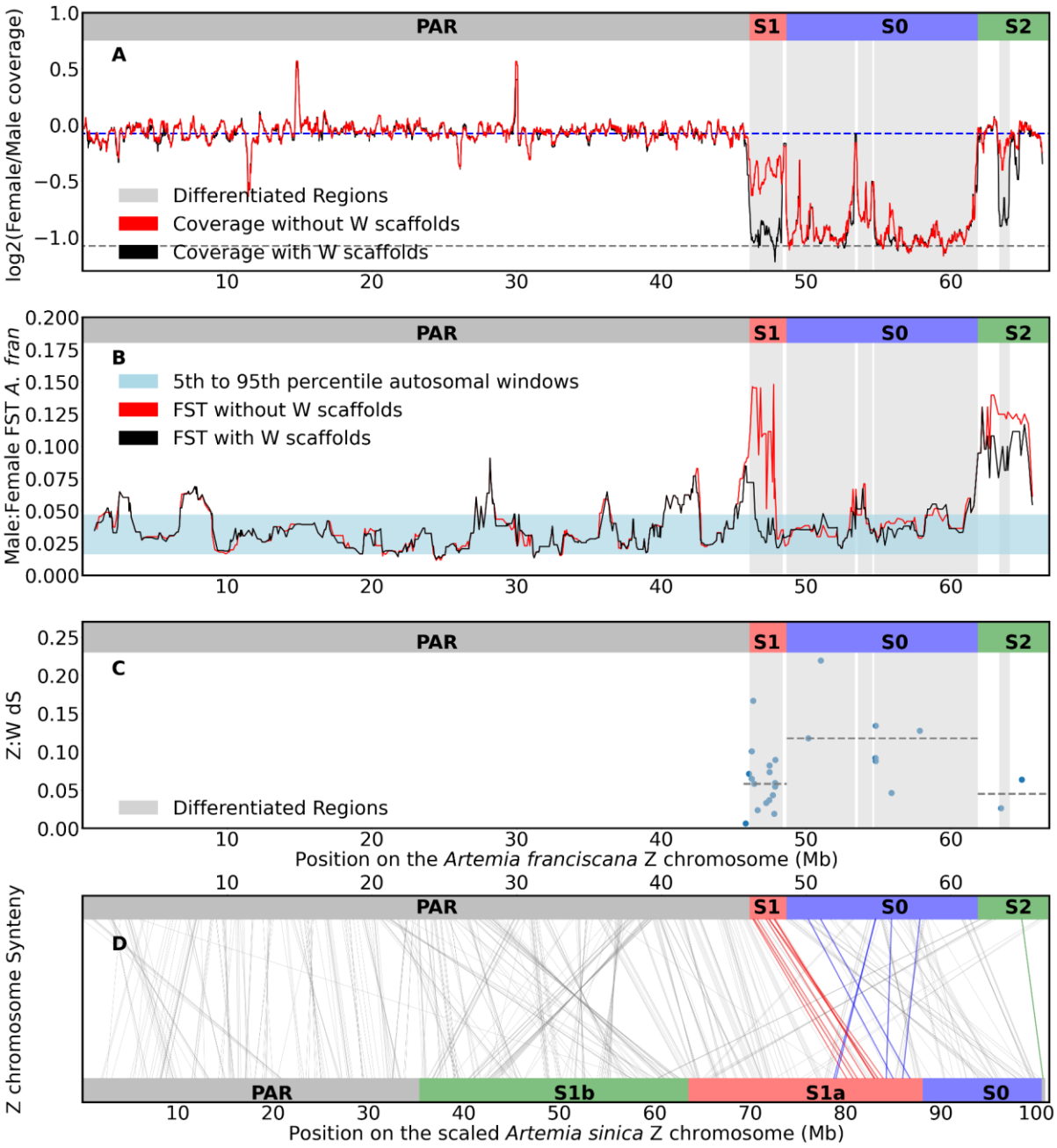


Figure 2: Evolutionary strata of the ZW pair. A) Log2(Female/Male) coverage patterns of the Z chromosome. The black and red lines depict the rolling median for 30 consecutive (10000 bp) windows estimated with and without including the W-scaffolds in the analysis. The gray-shading highlights the differentiated region of the Z chromosomes. The blue dashed line is the autosomal median for the analysis without W scaffolds, and the dashed gray line is at (median -1). B) The black and red lines are the rolling median Male:Female F_{ST} per gene for 10 genes estimated with and without including the W scaffolds, and the shaded light blue 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 gray 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 (light blue) S1 (light red) and S2 (light green). 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 and gonads (Figure 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 & Walters 2017). So far, Lepidoptera (moths and butterflies) have been the only clear exception to this rule (Gu & 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.

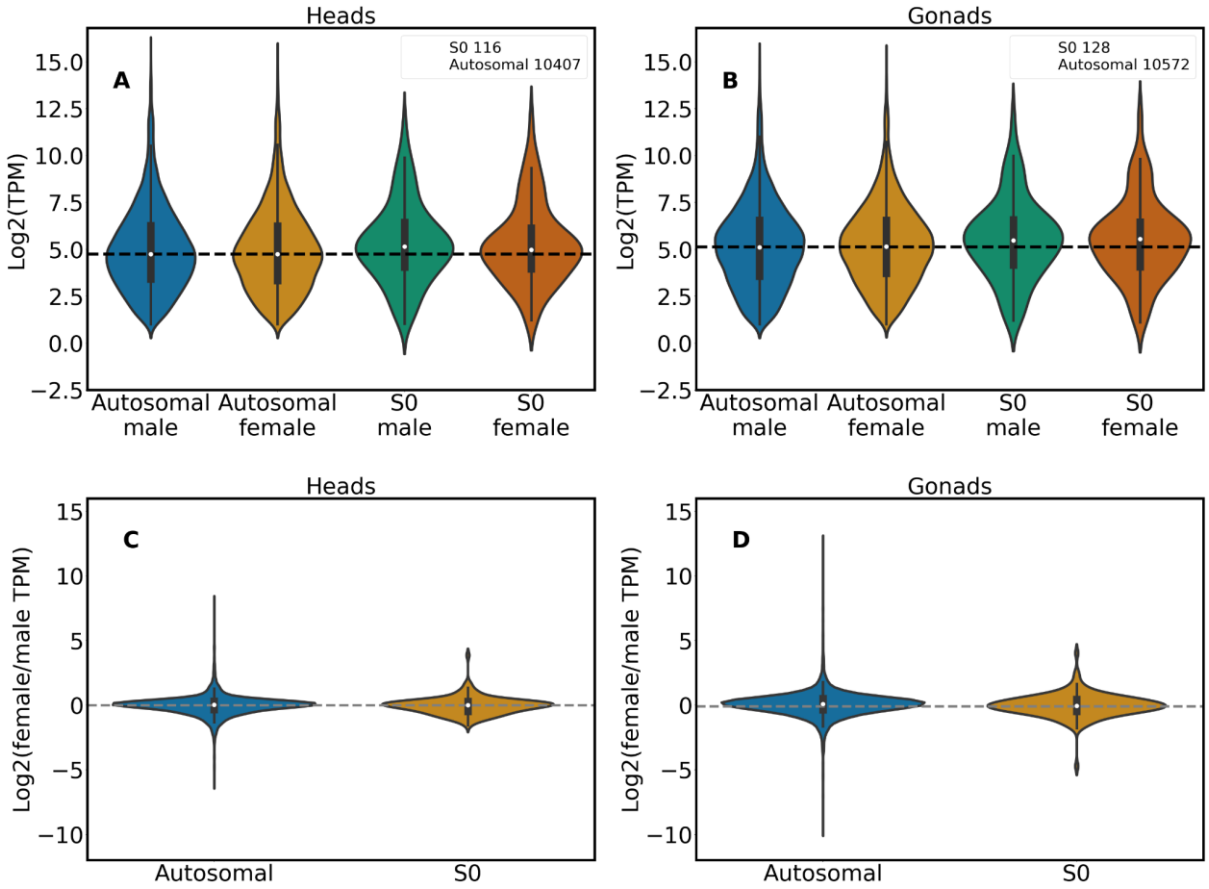


Figure 3: Dosage Compensation. A) The Log2 of the expression of Autosomal and the Z differentiated region genes in male and female heads. The legend shows the number of genes used in the analysis (TPM >= 0.5 in males and females). B) The Log2 expression of Autosomal and the Z differentiated region (S0) genes in gonads. The legend shows the number of genes used in the analysis (TPM >= 0.5 in males and females). C) The Log2 of female/male expression for the autosomal and Z differentiated genes (TPM >= 0.5 in males and females) in Heads D) The Log2 of female/male expression for the autosomal and Z differentiated genes (TPM >= 0.5 in males and females) in gonads.

Finally, the absence of recombination between sex chromosomes often leads to the accumulation of transposable elements and other repetitive sequences on the Y or W chromosome (Dechaud et al. 2019). Approximately two thirds of the genome in *A. franciscana* are covered by repetitive elements (Supplementary Table 2). 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 2). Overall, even this region of the W only shows a modest enrichment in repeats compared with the autosomes (67 versus 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 non-recombining part of the W in this clade (or limited power to resolve repeats in the assembly), another possibility is that the ancestral non-recombining 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 two unmated females from the great salt lake, purchased from Sanders (Utah, USA), 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 k-mers 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 two 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 Figure 3). RagTag (v2.1.0, Alonge et al. 2022) was used to implement those merges. Redundans (v0.14a, Pryszcz & Gabaldón 2016) was then used with long insert mate-pairs (SRR6980924, 8MB) to further scaffold the assembly.

Scaffolding using the published linkage map

The published *Artemia 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 two 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 linkage group 6 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 two 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 re-scaffold 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. 2023). The assessment of the assembly completeness was done using BUSCO (v5.2.2), with the arthropoda_odb10 dataset (Manni et al. 2021), and blobtools was used to assess and visualize the level of contamination and genome statistics (Laetsch & 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 & Salzberg 2012). The resulting SAM files were filtered for unique alignments using (grep -vw "XS:i"), and the coverage was estimated for windows of 10000 bp using soap.coverage (version 2.7.7, https://github.com/gigascience/bgi-soap2/tree/master/tools/soap.coverage/2.7.7).

In the F_{ST} 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 two 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 (Genome Annotation section) to estimate the F_{ST} per Gene (Kofler et al. 2011).

The coordinates for the differentiated region (Grey-shaded areas figure 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 [46085001 to 48385001],[48665001 to 53365001],[53585001 to 54575001],[54715001 to 61855001],[63345001 to 64075001], shaded in gray in Figure 2 A, B and 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 Z-W 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 *Artemia franciscana* genome assembly to generate a *de novo* library of repeat families. The sequences of unknown transposable elements (TE) were classified using DeepTE (Yan et al. 2020) with the options ‘*-sp M -m M”.* Subsequently, Class II transposable elements (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 of that DeepTE and used for the characterization of repeat content and masking of the genome by RepeatMasker (v4.1.5, Tarailo-Graovac & 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 augustus hint 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, incomplete genes without start and/or stop codons, and filtered out those with less than 100 bp length of Open Reading Frames. Protein sequences and coding sequences were extracted using getAnnoFastaFromJoingenes.py and their completeness were assessed with BUSCO (v5.2.2).

An annotation was produced for the *Artemia 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 & Kong 2019) (protein target and query) and reciprocal best hits were found using a customized python script and used for producing Figure 2 D.

Expression Analysis

Illumina RNA reads of two 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 & Dewey 2011). Transcript abundances of genes (in TPM) were estimated using rsem-calculate-expression in RSEM with options "--paired-end --alignments --estimate-rspd --strandedness no". We then used NormalyzerDE (Willforss et al. 2019) to perform quantile normalization across samples for each tissue separately. We applied different cut-offs of TPM >= 0.0, TPM >= 0.5, and TPM >= 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 (Figure 2 A,B,C, and D, Supplementary Figure 8 A and B, Supplementary Figure 9 A and B).

Data and Code Availability

The assembly pipeline and 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 NCBI short read archive (bioproject number PRJNA1017357). The annotation and the final assembly can be accessed through [will be added before publication]. All genomic and transcriptomic samples used for this study and the steps of the analysis they were used in are described in Supplementary table 3.

References

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

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

Alonge M et al. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 23:258. doi: 10.1186/s13059-022-02823-7.

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

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

Bowen ST. 1963. THE GENETICS OF *ARTEMIA SALINA*. III. EFFECTS OF X-IRRADIATION AND OF FREEZING UPON CYSTS. Biol. Bull. 125:431–440. doi: 10.2307/1539357.

Boyer L et al. 2023. Asexual male production by ZW recombination in *Artemia parthenogenetica*. Evolution. 77:1–12. doi: 10.1093/evolut/qpac008.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics Bioinforma. 2:lqaa026. doi: 10.1093/nargab/lqaa026.

Bushnell B. 2014. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States) https://www.osti.gov/biblio/1241166 (Accessed September 4, 2023).

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

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

Catchen J, Amores A, Bassham S. 2020. Chromonomer: A Tool Set for Repairing and Enhancing Assembled Genomes Through Integration of Genetic Maps and Conserved Synteny. G3 GenesGenomesGenetics. 10:4115–4128. doi: 10.1534/g3.120.401485.

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

Coombe L et al. 2021. LongStitch: high-quality genome assembly correction and scaffolding using long reads. BMC Bioinformatics. 22:534. doi: 10.1186/s12859-021-04451-7.

Criel G, Macrae T. 2002. Reproductive biology of *Artemia*. In: *Artemia*: Basic and applied Biology. Kluwer Academic pp. 39–128. http://hdl.handle.net/1854/LU-150279 (Accessed September 4, 2023).

Czech L, Spence JP, Expósito-Alonso M. 2023. grenedalf: population genetic statistics for the next generation of pool sequencing. doi: 10.48550/ARXIV.2306.11622.

Dainat J et al. 2023. NBISweden/AGAT: AGAT-v1.2.0. doi: 10.5281/ZENODO.8178877.

De Vos S et al. 2013. A first AFLP-Based Genetic Linkage Map for Brine Shrimp *Artemia franciscana* and Its Application in Mapping the Sex Locus Liu, Z, editor. PLoS ONE. 8:e57585. doi: 10.1371/journal.pone.0057585.

De Vos S et al. 2021. The genome of the extremophile *Artemia* provides insight into strategies to cope with extreme environments. BMC Genomics. 22:635. doi: 10.1186/s12864-021-07937-z.

Dechaud C, Volff J-N, Schartl M, Naville M. 2019. Sex and the TEs: transposable elements in sexual development and function in animals. Mob. DNA. 10:1–15. doi: 10.1186/s13100-019-0185-0.

Dobin A et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21. doi:

10.1093/bioinformatics/bts635.
Eimanifar A, Van Stappen G, Wink M. 2015. Geographical distribution and evolutionary divergence times of Asian populations of the brine shrimp *Artemia* (Crustacea, Anostraca): Divergence times of Asian *A rtemia*. Zool. J. Linn. Soc. 174:447–458. doi: 10.1111/zoj.12242.
Elkrewi M et al. 2022. ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp Dyer, K, editor. Genetics. 222:iyac123. doi: 10.1093/genetics/iyac123.
Elkrewi M, Moldovan MA, Picard MAL, Vicoso B. 2021. Schistosome W-Linked Genes Inform Temporal Dynamics of Sex Chromosome Evolution and Suggest Candidate for Sex Determination Wilson, M, editor. Mol. Biol. Evol. 38:5345–5358. doi: 10.1093/molbev/msab178.
Flynn JM et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117:9451–9457. doi: 10.1073/pnas.1921046117.
Formenti G et al. 2022. Merfin: improved variant filtering, assembly evaluation and polishing via k-mer validation. Nat. Methods. 19:696–704. doi: 10.1038/s41592-022-01445-y.
Gu L, Walters JR. 2017. Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details. Genome Biol. Evol. 9:2461–2476. doi: 10.1093/gbe/evx154.
Guan D et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies Valencia, A, editor. Bioinformatics. 36:2896–2898. doi: 10.1093/bioinformatics/btaa025.
Han X, Ren Y, Ouyang X, Zhang B, Sui L. 2021. Construction of a high-density genetic linkage map and QTL mapping for sex and growth traits in *Artemia franciscana*. Aquaculture. 540:736692. doi: 10.1016/j.aquaculture.2021.736692.
Handley L-JL, Ceplitis H, Ellegren H. 2004. Evolutionary Strata on the Chicken Z Chromosome: Implications for Sex Chromosome Evolution. Genetics. 167:367–376. doi: 10.1534/genetics.167.1.367.
Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-Genome Annotation with BRAKER. In: Gene Prediction. Kollmar, M, editor. Methods in Molecular Biology Vol. 1962 Springer New York: New York, NY pp. 65–95. doi: 10.1007/978-1-4939-9173-0_5.
Hu J et al. 2023. *NextPolish2:a repeat-aware polishing tool for genomes assembled using HiFi long reads*. Bioinformatics doi: 10.1101/2023.04.26.538352.
Huylmans AK, Toups MA, Macon A, Gammerdinger WJ, Vicoso B. 2019. Sex-Biased Gene Expression and Dosage Compensation on the *Artemia franciscana* Z-Chromosome Mank, J, editor. Genome Biol. Evol. 11:1033–1044. doi: 10.1093/gbe/evz053.
Jo E et al. 2021. Whole genome survey and microsatellite motif identification of *Artemia franciscana*. Biosci. Rep. 41:BSR20203868. doi: 10.1042/BSR20203868.
Kofler R et al. 2011. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals Kayser, M, editor. PLoS ONE. 6:e15925. doi: 10.1371/journal.pone.0015925.
Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. F1000Research. 6:1287. doi: 10.12688/f1000research.12232.1.
Lahn BT, Page DC. 1999. Four Evolutionary Strata on the Human X Chromosome. Science. 286:964–967. doi: 10.1126/science.286.5441.964.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359. doi: 10.1038/nmeth.1923.
Lavens P, Sorgeloos P. 2000. The history, present status and prospects of the availability of *Artemia* cysts for aquaculture. Aquaculture. 181:397–403. doi: 10.1016/S0044-8486(99)00233-1.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323. doi: 10.1186/1471-2105-12-323.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. doi: 10.48550/ARXIV.1303.3997.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences Birol, I, editor. Bioinformatics. 34:3094–3100. doi: 10.1093/bioinformatics/bty191.
Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics.

25:2078–2079. doi: 10.1093/bioinformatics/btp352.
Lovell JT et al. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife. 11:e78526. doi: 10.7554/eLife.78526.
Maniatsi S, Kappas I, Baxevanis A, Farmaki T, Abatzopoulos T. 2009. Sharp Phylogeographic Breaks and Patterns of Genealogical Concordance in the Brine Shrimp *Artemia franciscana*. Int. J. Mol. Sci. 10:5455–5470. doi: 10.3390/ijms10125455.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes Kelley, J, editor. Mol. Biol. Evol. 38:4647–4654. doi: 10.1093/molbev/msab199.
Mc Cartney AM et al. 2022. Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies. Nat. Methods. 19:687–695. doi: 10.1038/s41592-022-01440-3.
Nunes BS, Carvalho FD, Guilhermino LM, Van Stappen G. 2006. Use of the genus *Artemia* in ecotoxicity testing. Environ. Pollut. 144:453–462. doi: 10.1016/j.envpol.2005.12.037.
Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 44:e113–e113. doi: 10.1093/nar/gkw294.
Rajabi S, Ramazani A, Hamidi M, Naji T. 2015. *Artemia salina* as a model organism in toxicity assessment of nanoparticles. DARU J. Pharm. Sci. 23:20. doi: 10.1186/s40199-015-0105-x.
Song L, Shankar DS, Florea L. 2016. Rascaf: Improving Genome Assembly with RNA Sequencing Data. Plant Genome. 9:plantgenome2016.03.0027. doi: 10.3835/plantgenome2016.03.0027.
Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Curr. Protoc. Bioinforma. 25. doi: 10.1002/0471250953.bi0410s25.
Van Stappen G et al. 2020. Review on integrated production of the brine shrimp *Artemia* in solar salt ponds. Rev. Aquac. 12:1054–1071. doi: 10.1111/raq.12371.
Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27:737–746. doi: 10.1101/gr.214270.116.
Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics Proteomics Bioinformatics. 8:77–80. doi: 10.1016/S1672-0229(10)60008-3.
Wang M, Kong L. 2019. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinformatics. 20:1–4. doi: 10.1186/s12859-019-2597-8.
Willforss J, Chawade A, Levander F. 2019. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. J. Proteome Res. 18:732–740. doi: 10.1021/acs.jproteome.8b00523.
Xu M et al. 2020. TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience. 9:giaa094. doi: 10.1093/gigascience/giaa094.
Yan H, Bombarely A, Li S. 2020. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. Bioinformatics. 36:4269–4275. doi: 10.1093/bioinformatics/btaa519.

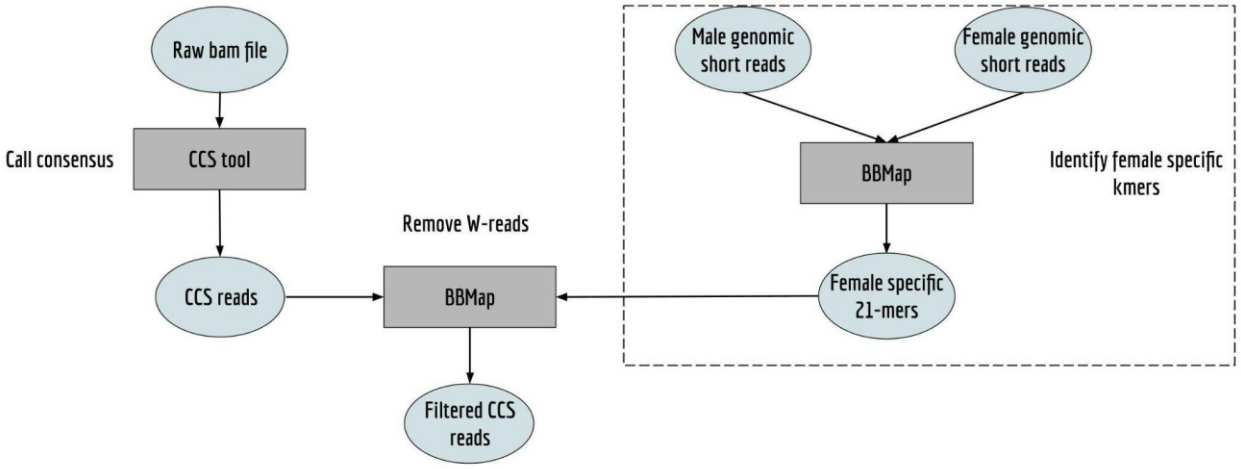
Supplementary Material

Supplementary Datasets

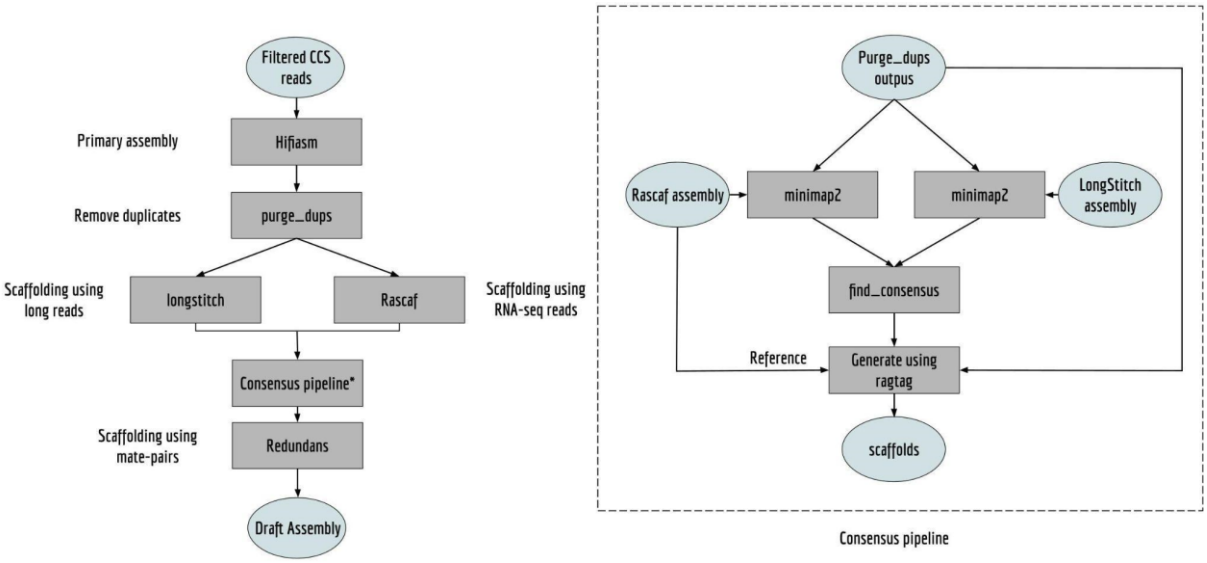
Supplementary Dataset 1: *Artemia franciscana* genome assembly and annotation
Supplementary Dataset 2: *Artemia sinica* annotation
Supplementary Dataset 3: Male and Female genomic coverage
Supplementary Dataset 4: Male:female F_{ST} per gene
Supplementary Dataset 5: Normalized gene expression
Supplementary 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

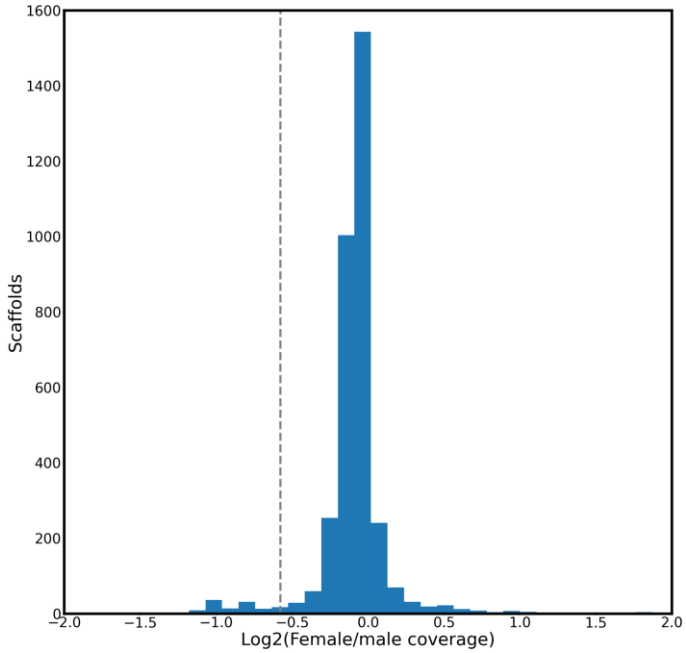
Supplementary Figures



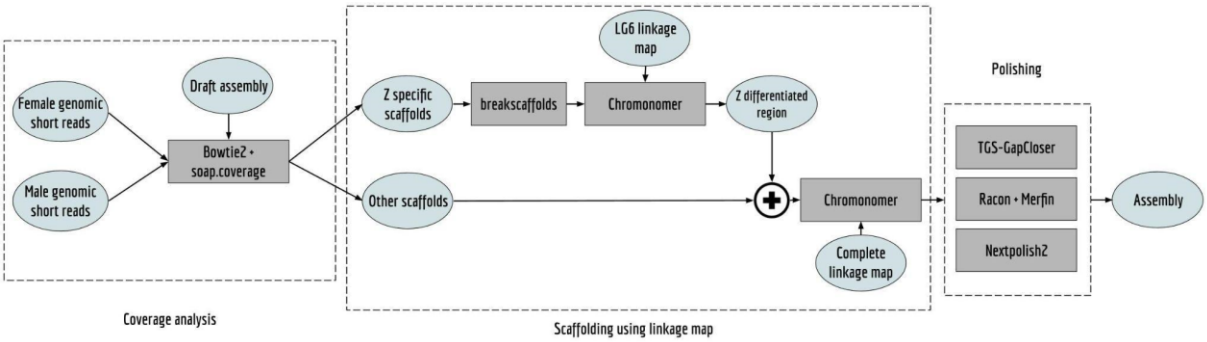
Supplementary Figure 1: the pipeline from generating the consensus reads, and Identifying and removing putative W-reads using female specific k-mers identified using a subtraction approach with male and female genomic short reads.



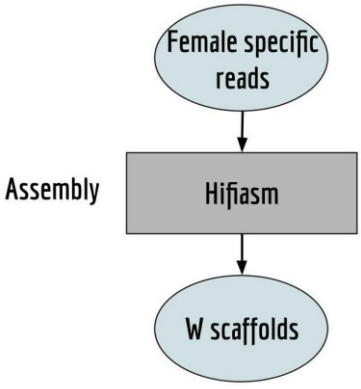
Supplementary Figure 2: Assembling the genome and scaffolding with genomic long reads and RNA-seq reads



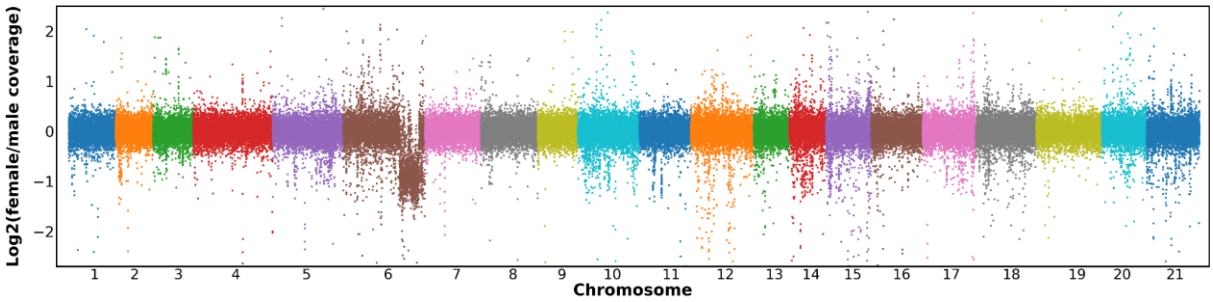
Supplementary Figure 3: Coverage analysis used to identify the scaffolds belonging to the differentiated region of the Z chromosome. The gray-dashed line represents the threshold used: (Log2(female/male) < median(Log2(female/male)-0.5).



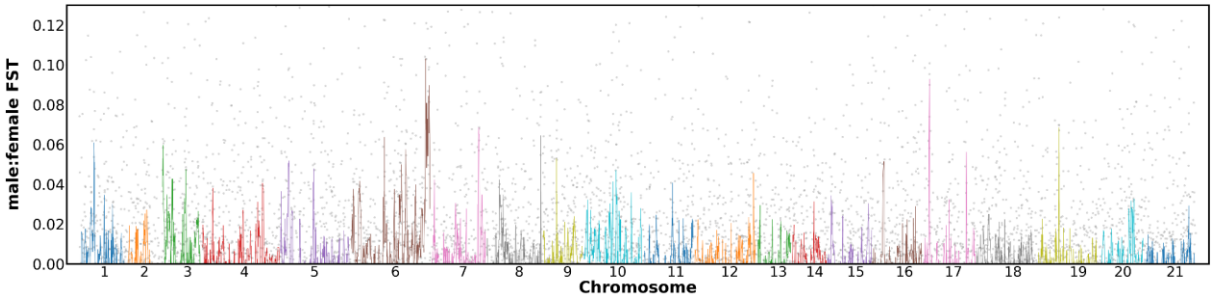
Supplementary Figure 4: The pipeline for scaffolding the genome using the published linkage map



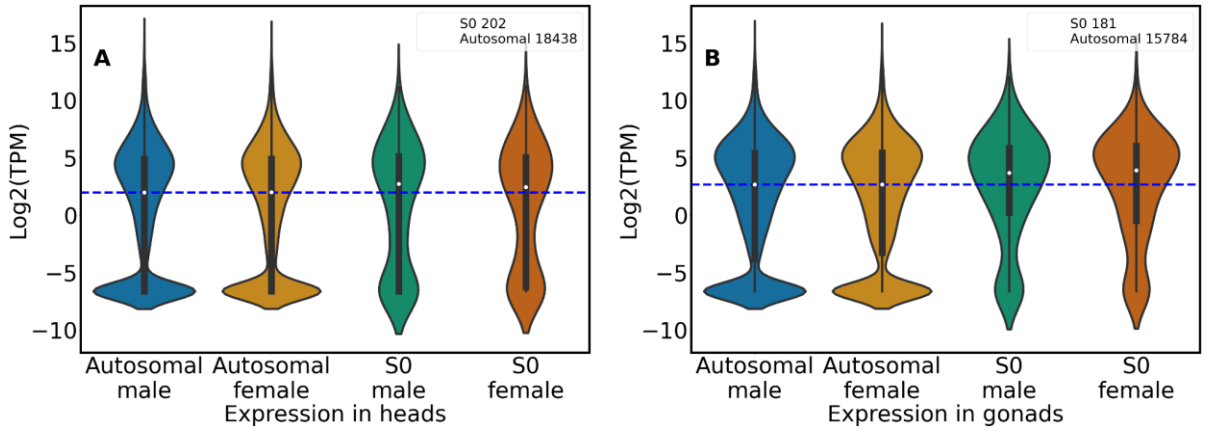
Supplementary Figure 5: Assembly of W scaffolds



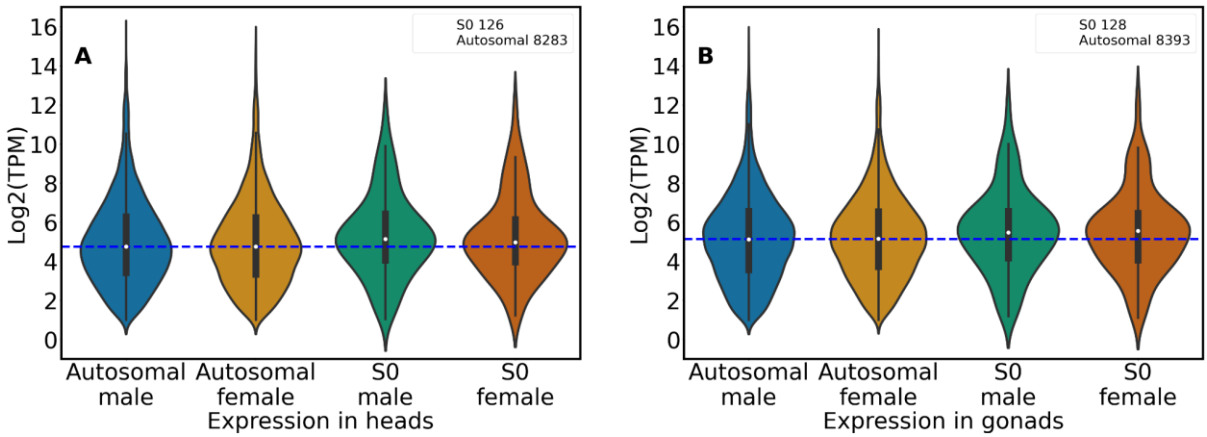
Supplementary Figure 6: Log2(Female/Male) coverage patterns across the whole genome (W-scaffolds included).



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://docs.google.com/spreadsheets/d/1X_Q8FI_BxrNtsJ7YFpmsjcU-C718pCk0xlMxY6wYkhU/edit?usp=sharing

CHAPTER 3

A novel regulatory mechanism for the conserved sex determination gene Doublesex in Artemia brine shrimp

Chapter 3: A novel regulatory mechanism for the conserved sex determination gene Doublesex in Artemia brine shrimp

Authors:
Marwan Elkrewi[1]@, Ruben Olbrechts[1], Ariana Macon[1], Beatriz Vicoso[1]@
@corresponding authors
[1] Institute of Science and Technology Austria (ISTA), Klosterneuburg 3400, Austria
Key words:
Sex determination, Sexual dimorphism, Doublesex, Feminizer, C2H2 zinc finger proteins, lncRNA

Abstract

The regulation of sexually dimorphic traits often involves a complex network of interactors that vary in the extent of their conservation across short and long evolutionary times. The upstream regulators of this cascade tend to undergo rapid turnovers, while the downstream regulators show more conservation even across distant species. Despite the high rates of turnover of the upstream regulators, evidence suggests that some genes are more likely to become sex determination genes, with the majority of the identified master regulators in vertebrates belonging to the TGF-β superfamily or to certain families of transcription factors (DMRT and SOX)(1,2). Members of the DMRT family of transcription factors are among the most conserved downstream regulators of primary and secondary sexual characteristics in a broad range of species, with many instances of secondary recruitment to the role of the master determiner. In most Arthropods, the control of the sex-specific functions of doublesex (a member of the DMRT family) is mediated through alternative-splicing, with either a male and female functional isoforms, or a functional isoform in males that initiates testicular development, and the lack of it initiating ovarian development in females. In this paper, we performed an extensive search for known and unknown genes involved in the sex determination and differentiation cascades in *Artemia*, a branchiopod crustacean, using a combination of newly generated stranded RNAseq data of nauplii (the first larval stage) and the previously published genomic and transcriptomic datasets. We report a novel mechanism for the regulation of the sex-specific expression of Dsx in *Artemia* brine shrimp, by an overlapping antisense lncRNA in females, which we refer to as SUM (Suppressor of Masculinization). We confirmed the expression patterns of DSX and SUM using standard PCR and real-time quantitative PCR across various developmental stages to paint a picture of their expression dynamics. We have also identified a putative upstream regulator of SUM (FEM) and characterized its potential binding site in the promoter region. Additionally, we have perturbed the expression/translation of the transcription factor in primary cell culture generated from *Artemia parthenogenetica* cysts, using dsRNA. The manipulation resulted in a higher expression of DSX in the female cells, suggesting a repressive role for the feminizer gene.

Introduction

In species with males and females, sex is determined either through genetic, environmental, or cytoplasmic factors (3). In systems with genetic sex determination (GSD), the chromosomal sex of the embryo is determined at conception. However, the timing in which the sex determination pathway is activated (sex determination window) and the genes involved vary widely between different species (4). In humans, for instance, sex is determined by a Y-linked gene, SRY, which is first expressed six weeks after conception in gonadal somatic cells, triggering their differentiation into testicular supporting cells that facilitate the differentiation of other cells into their sex-specific fates. In females, the lack of SRY leads to the gonadal somatic cell's differentiation into ovarian support cells (5). In *Drosophila*, on the other hand,

through the activity of counter genes, the ratio of X chromosomes to Autosomes controls the activation of the master switch, Sex-lethal (Sxl). The expression of Sxl is detectable in *Drosophila* female embryos at mitotic stage 13, not long after the zygotic genome activation and before the completion of cellularization (6). Studying sex determination across the tree of life has uncovered various mechanisms and architectures, but within this extreme diversity, various recurrent themes are also observed.

One of the the most striking commonalities in the sex determination and differentiation cascades is the involvement of genes from the Dmrt family of transcription factors (7). Across the tree of life, Dmrt genes are identifiable with their conserved zinc-finger DNA binding domain (DM domain), and they seem to have ancestrally acted as tissue-specific developmental regulators that promote/repress male or female gonadal development (7). In many species, Dmrt genes have acquired other sex-specific functions through duplications/neofunctionalization, including becoming the master sex determination gene, as in the african clawed frog (*Xenopus laevis*) and medaka *(Oryzias latipes*) (3), or controlling secondary sexual traits/characteristics, as in the large mandibles in male beetles (8).

Comparative studies of the DMRT gene Doublesex (Dsx) and its functions in arthropod sexual differentiation have shed light on its evolutionary history and its transition from being essential only for male differentiation to acquiring its additional role in female development in holometabolous insects (9). This was likely possible due to a duplication event in the common ancestor of branchiopoda and hexapoda and evolution of sex-specific splicing of Dsx, a mechanism that has only been reported in arthropods (9,10). The female isoform became essential for morphology later in holometabolous insects as a result of the extension of the C-terminal domain.

While our understanding of the sex-specific regulation of Doublesex in hexapods is very extensive, the picture is less clear in other arthropods, where only a few cases have been studied and the general understanding of sex determination and differentiation mechanisms is still lacking. In terms of the crustaceans surveyed so far, multiple mechanisms have been reported for the regulation of Dsx. For instance, in the Malacostracan, Eriocheir sinensis, Dsx does not bind to DNA directly; instead, it forms a heterodimer with another Dmrt protein and they bind to the promoter of Insulin-like androgenic hormone (IAG), which stimulates male development (11). In the Branchiopod, *Daphnia*, Dsx is regulated by a transcription factor (Vrille) and doublesex1 alpha promoter-associated long non-coding RNA (DAPALR). They function downstream of a hormonal photoperiod-sensitive cascade (12). Despite the extreme diversity of sex determination systems in crustaceans, the species sampled so far suggest that Dsx is still at the center of the sex determination and differentiation cascades. Therefore, understanding the various ways in which it is regulated will lead to a better understanding of the evolution of sex determination in crustaceans overall, where very little is currently known (12).

Crustaceans of the genus *Artemia* have ZW sex chromosomes, which are shared across the clade, with varying degrees of differentiation (13–15). All the species have a highly differentiated region termed S0, which has been completely lost from the W. Younger non-recombining regions have been identified in *Artemia sinica* and *Artemia franciscana*, along with large pseudoautosomal regions (16,17). The identified gene content on the W seems to be largely from the younger strata, and no genes showed patterns consistent with being ancestrally on the W chromosome. This could be partly due to the limitations of the approaches used and the lack of a complete assembly of the W chromosome. Multiple Dsx isoforms have been reported, and three recent papers suggested that male and female isoforms are generated through alternative splicing (18–20). The doublesex gene in *Artemia* is autosomal, and to date, no upstream regulator for its sex specific pattern has been identified. The lack of RNAseq data from the earlier stages of development leaves two possibilities open: a dominant W gene that is only transiently expressed, or a dosage sensitive mechanism involving genes on the differentiated region of the Z chromosome before establishing dosage compensation (No consistently male biased gene on the S0 has been identified). We sought to fill the gap in our understanding of sex determination in *Artemia* using RNAseq data from single individuals at an early developmental time point from two distant species. We reasoned

that the number of sex biased genes in early development is small compared to the later stages (21), and that if the sex determination genes are conserved, the shared patterns of sex biased gene expression could be used to narrow down the likely candidates.

Results

Patterns of sex-biased gene expression across development

As sex determination genes are generally expressed early in development, we hypothesized that if they are conserved across the *Artemia* genus, comparing the expression patterns in early development between two the distantly related species of *Artemia franciscana* and *Artemia sinica* (diverged 30 million years ago) could help identify promising candidates. Therefore, we performed single nauplii RNAseq on 14 individuals from each species (frozen within 1 day of their birth). We mapped the reads to their respective genomes using Hisat2 (22) and identified the sex of the nauplii using PCA (PLINK) on the differentiated region of the Z-chromosome, along with the number of reads that mapped with no mismatches to previously published W-transcripts (16), Supplementary Figure 1). Our samples had 7 males and 7 females in A. franciscana, and 8 males and 6 females in A. sinica.

As the previously published annotations were made using only adult RNAseq data, they may be missing genes that are expressed only expressed in earlier stages; therefore, we re-annotated the genomes of the two species by combining the published adult data (15) and the newly generated nauplii data using StringTie2 (23). As the *Artemia sinica* assembly does not have annotated W-scaffolds, we used a kmer approach to assemble W-transcripts from both adults and nauplii (16,24). We added the assembled transcripts to the genome before annotation. The *Artemia franciscana* genome (17) includes W-scaffolds, so no further analysis was required. With the two genomes annotated, we performed differential expression analysis using DEseq2 (25) in each species separately for both nauplii and adults, and we compared the results using reciprocal best hits of the longest isoform of each gene. As expected, while adults show a large number of genes with concordant patterns of sex-bias, nauplii have only a few genes shared between the two species (Figure 1 A and B).

Four genes had consistent patterns between the two species (adjusted p-value <0.05): two slightly male biased genes and two very female-biased genes. One of the two female biased genes in Nauplii is a C2H2 transcription factor (depicted in blue), and the other is the previously identified female isoform of Dsx (depicted in orange) (20). The male isoform of Dsx (depicted in green) is male biased in the *Artemia sinica* nauplii samples, but not yet expressed in the *Artemia franciscana* nauplii samples. This could possibly be due to the samples of the two species having been collected at different timepoints within the first 24 hours. In adults, Dsx shows extreme male bias in both species, consistent with the previous reports. The female isoform of DSX and the other female biased gene continue to show female biased expression patterns in adults as well (Figure 1 B). We confirmed the patterns using qPCR of adult male and female samples (Figure 1 C, D, and E).

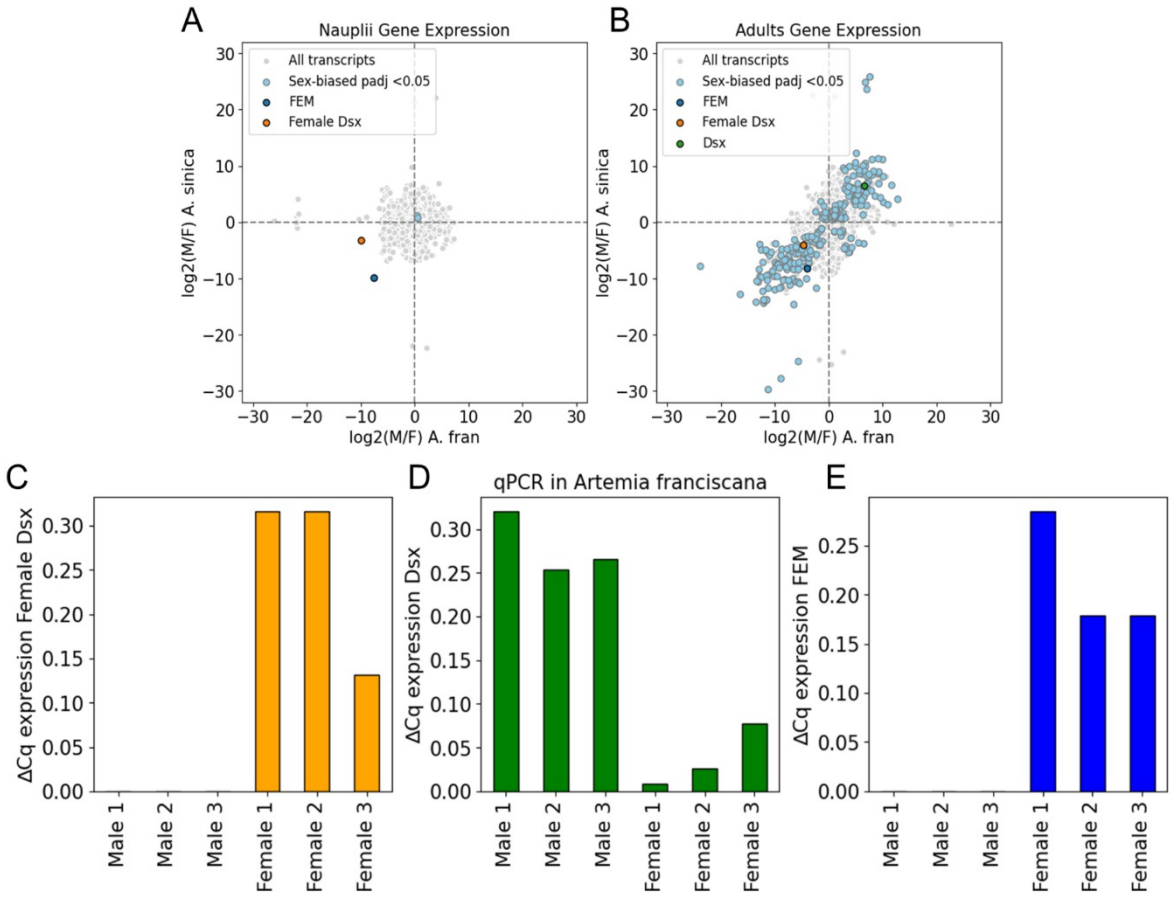


Figure 1: Differential expression analysis results. A) Nauplii log2(Male/Female expression) in A. sinica vs log2(Male/Female expression) in A. franciscana. B) Adults log2(Male/Female expression) in A. sinica vs log2(Male/Female expression) in A. franciscana. C) qPCR of the female isoform of DSX in A. franciscana colony individuals (3 adult males and 3 adult females). D) qPCR of the male isoform of DSX in A. franciscana colony individuals (3 adult males and 3 adult females). E) qPCR of the C2H2 zinc finger in A. franciscana colony individuals (3 adult males and 3 adult females).

Dsx is regulated by the expression of an antisense overlapping gene in females, and not by alternative splicing

To confirm and further characterize the reported patterns of alternative splicing of Dsx in *Artemia*, we mapped the published male and female isoforms to transcript models assembled using our RNAseq data and the chromosome-level genome assembly. The two isoforms mapped to two different transcripts in the same genomic location, but on opposite strands. The alignments on the genome viewer (IGV) (26) for the adult male and female stranded libraries confirmed that the female and male reads align in opposite directions, with the transcript in the sense direction encoding for Dsx and expressed only in males. The overlapping transcript in females had various isoforms (26 in *Artemia sinica* and 45 in *Artemia franciscana*), with many overlapping Dsx intronic regions. Figure 2 A shows the number of reads that map in the sense (defined relative to Dsx) and antisense direction in the region surrounding Dsx and the gene models at the bottom. If antisense transcripts are being produced in the two sexes, these should have poly-A tails at opposite 3'ends, and this can be further used to verify antisense expression experimentally. As the RNAseq data lacks polyA tails, we used the published single-nucleus RNAseq data from the female reproductive system (27) to identify the poly-A tail for the antisense transcript using polyApipe (https://github.com/MonashBioinformaticsPlatform/polyApipe). We used the predicted polyA tails to design reverse primers for the antisense gene containing 14 specific bases and a polyT tail (10 bp) along with a specific forward primer to confirm the strandedness. As expected, when running PCR on male and female cdna samples (generated from whole body total RNA

extraction), we see that the antisense transcript amplifies only in females (Figure 2 B). The overlapping gene was also present in the published long read RNAseq data (Supplementary figure 2). Given its putative role in suppressing Dsx, we tentatively named this novel gene Suppressor of Masculinization (SUM).

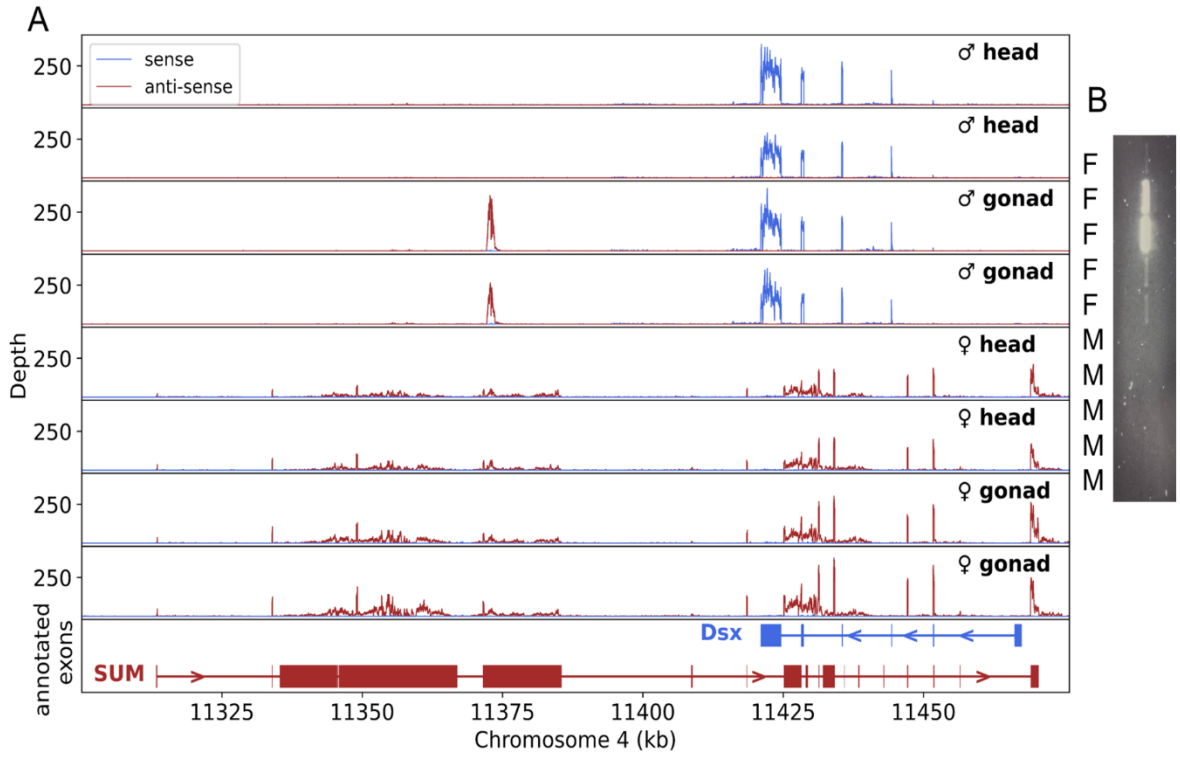


Figure 2: A) Read strandedness in the region surrounding Dsx. B) PCR of the antisense RNA using a specific forward primer and a reverse polyT primer.

The antisense lncRNA expression in females is conserved across the clade

We explored whether the overlapping gene expression in males and females was specific to *Artemia franciscana* or shared across the *Artemia* clade using published data from multiple sexual and asexual Asian species (28), including *Artemia sinica*, *Artemia sp. Kazakhstan*, and *Artemia parthenogenetica* populations from Aibi and Atanasovsko lakes. We observe the same patterns of expression, where the females only express the antisense transcript and the males express Dsx (Figure 3, A, B, C, D). We further extracted and sequenced RNA of rare males (produced at very low frequencies in asexual populations) from the atanasovsko population of *Artemia parthenogenetica*, and they show the same pattern as the males from the sexual species.

In order to characterize the nature of the antisense sequence, we used the tool lncDC (29), a machine learning based tool, to classify all the annotated transcripts into protein coding and long noncoding RNA. We trained the model separately using *Drosophila* and *Daphnia* coding and non-coding sequences. In *Artemia franciscana*, out of 45 isoforms of SUM, 44 were noncoding and 1 was coding. In *Artemia sinica*, out of 26 isoforms of SUM, 25 were noncoding and 1 was coding. The predicted coding isoforms (afran.41398.7 and asinica.62588.12) do not share any homology and do not overlap Dsx, suggesting that the non-coding transcripts may be the mediators of Dsx regulation. To further characterize whether the long non coding RNAs are conserved in sequence, position, and/or splicing patterns (30), we extracted the genomic regions surrounding Dsx from the repeat masked genomes of A. franciscana and A. sinica and used NCBI blast to align them. The regions are overall highly syntenic, with the coding regions of Dsx falling exclusively in highly conserved regions. While some of the exons of the overlapping antisense gene are in conserved regions, many others do not seem to be conserved. This is not surprising as long noncoding RNA do not evolve under the same

constraints as coding sequences (31), and in many cases the promoter region is the most conserved part of the long non-coding RNA (30).

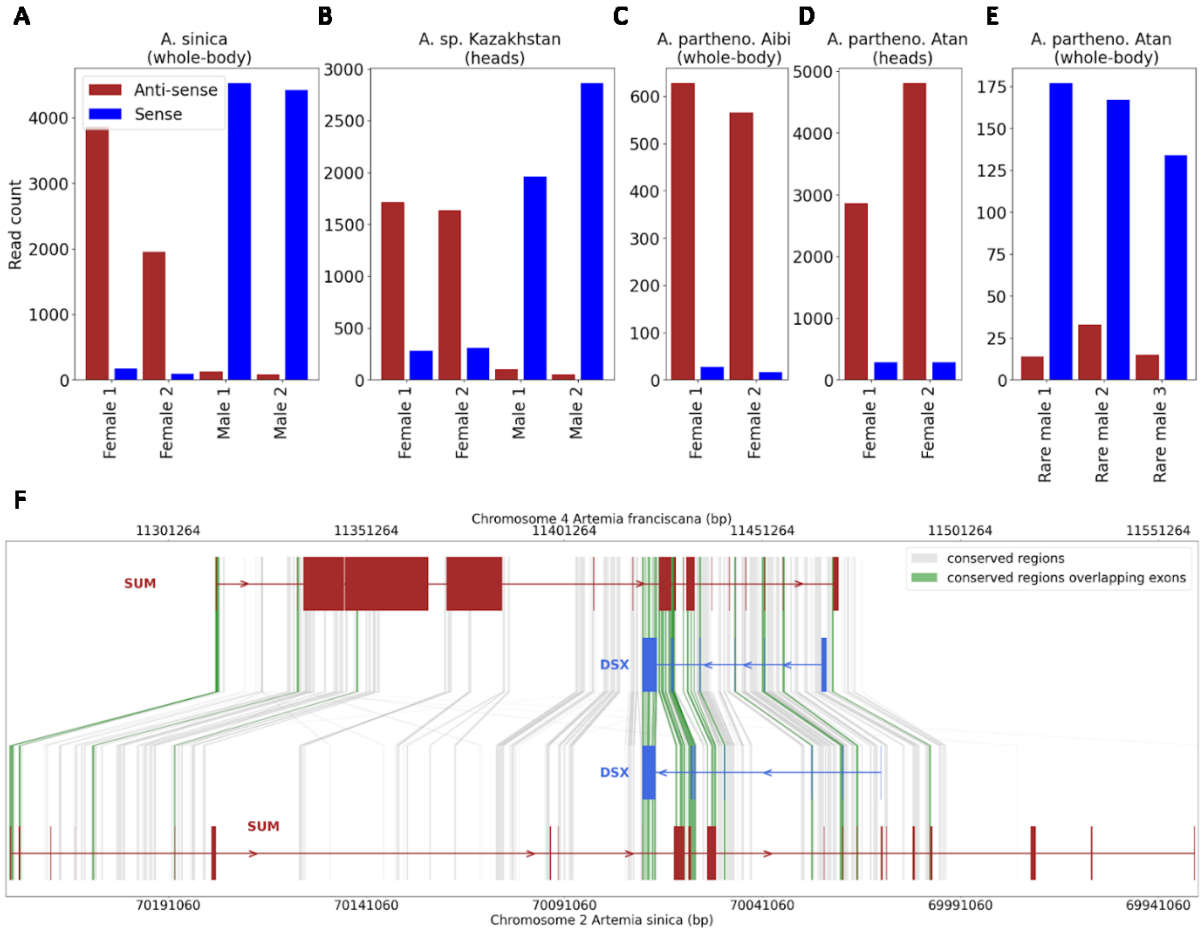


Figure 3: Conservation of the Dsx antisense regulatory mechanism across *Artemia species* A) *A. sinica* whole body RNAseq B) *A. sp. kazakhstan* heads RNAseq C) *A. parthenogenetica* (Aibi Lake) whole body RNAseq D) *A. parthenogenetica* (Atanasovsko) heads RNAseq E) *A. parthenogenetica* (Atanasovsko) rare males whole body RNAseq F) Sequence-level conservation in the region surrounding Dsx between *A. sinica* and *A. franciscana*.

Sex specific histone modifications suggest promoter competition and a potential link to dosage compensation

Histone modifications are known to play an essential role in the control and maintenance of sex specific programs. In *Artemia*, for instance, H4K16ac has been implicated in the upregulation of the S0 genes in females to match the male dosage; other active and repressive histone marks are also highly correlated with expression patterns (Vincent et. al, in prep). In order to explore the role of active and repressive histone modifications in the regulation of Dsx or the antisense overlapping gene, we used the published histone modifications data of *Artemia franciscana* male and female heads and gonads generated using CUT&Tag (32). We generated the alignment bam files as described in (Vincent et. al, in prep), and used GoPeaks (33) to call the peaks in each sample separately, and then we merged the peaks using the R script provided on the gitpage. After counting the number of reads overlapping the merged peaks in each sample, we performed differential peak accessibility analysis using DEseq2 (25) and looked at all the significant hits near the DSX region (Figure 4, Supplementary Figure 3). We identified sex specific peaks for H4K3me3, which marks active promoters near the TSS of DSX in males and SUM in females. The female H4K3me3 peak near the antisense gene TSS is much larger in magnitude than the male-specific H4K3me3 peak. As SUM is expressed earlier, has a stronger promoter and an exon overlapping the putative Dsx promoter, its transcription could be interfering with RNA pol II binding to the promoter of Dsx. The female anti-sense Dsx also has an H4K16ac peak near the promoter region, a histone mark which

modulates dosage compensation in this group. This could hint at a link between the expression of SUM and the dosage compensation pathway. Most of the other histone modifications show sex-specific peaks, including H3k27ac, which marks active enhancers. It is important to note that in some cases it is not straightforward to disentangle whether the presence of a histone modification is causal or the result of secondary acquisition in regions of highly accessible chromatin due to the action of chromatin remodelers and/or transcription factors. In the case of SUM, it is possible that the female specific histone-modifications are acquired through the interaction of a transcription factor and histone-modifying enzymes (34).

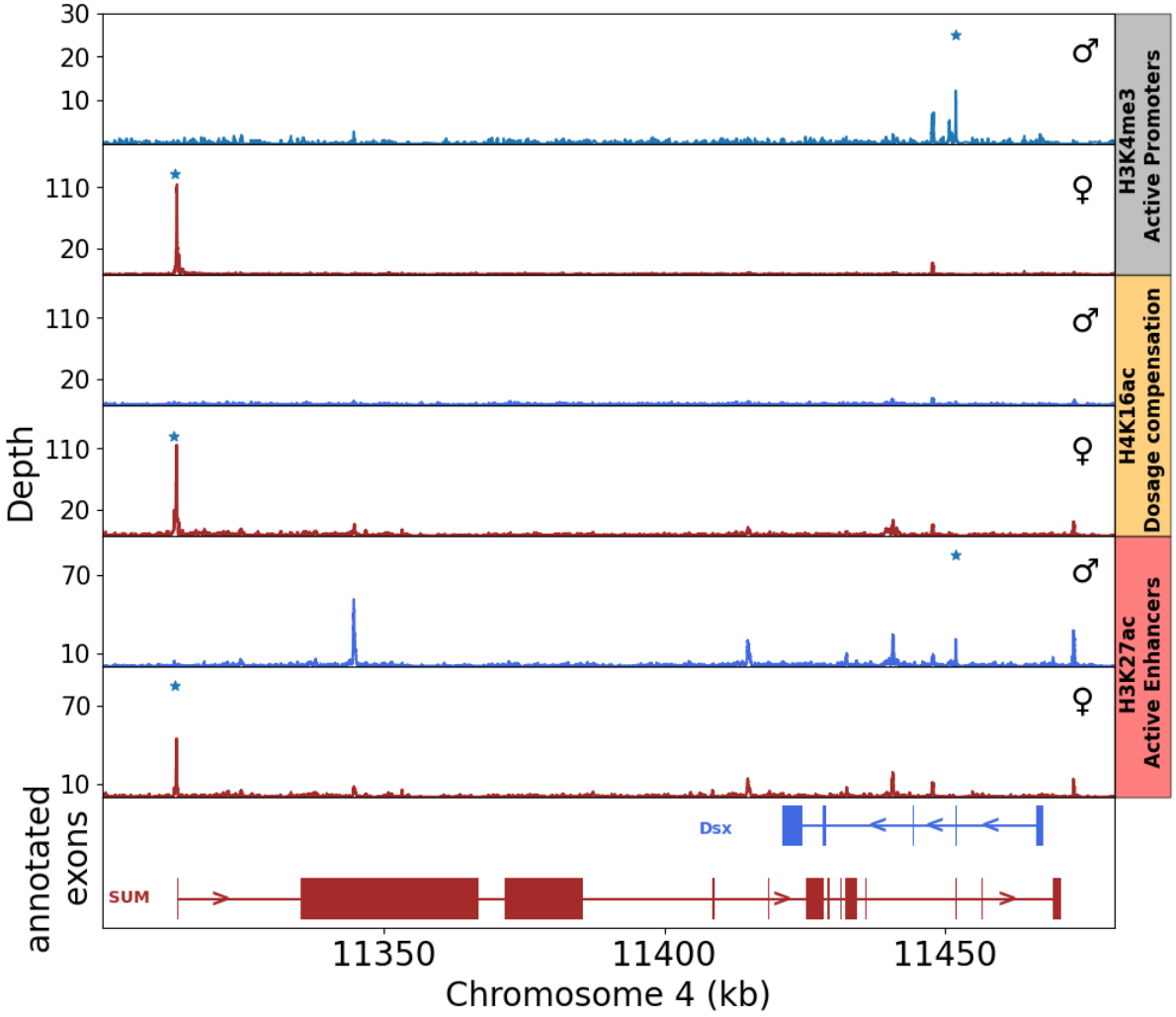


Figure 4: CUT&Tag data suggests a role for histone modifications in the sex-specific regulation of SUM and DSX.

The C2H2 zinc finger FEM could be regulating the expression of SUM and/or repressing DSX

Although the second female biased gene in our naupliar data was also autosomal, we found it highly plausible that it is part of the sex determination pathway due to its early expression (Supplementary Figure 4) and consistent female specificity throughout the lifespan. The results from lncDC (29) suggested that the gene is protein coding and the TransFacPred tool (35) classified it as a transcription factor. We used PROSITE (36) and NCBI conserved domain database (CDD) (37) to search the protein sequence for conserved domains. We found that the gene contains 8 C2H2 zinc fingers. C2H2 zinc fingers are one of the most abundant TF families in eukaryotes, with around 800 of them identified in humans (38). As the name suggests, C2H2 zinc fingers contain two cysteine and two histidine residues which form a tetrahedron with a zinc atom (39). The amino-acid sequence (12 amino acids in most cases) between the two 2 Cys and 2 His defines the unique nucleic acid sequence to which a zinc

finger binds. C2H2 TFs often contain tandem arrays of zinc fingers that tend to bind in promoter and enhancer regions (38). Zinc fingers are also known to act as transcriptional repressors as well (40,41).
The *Artemia* zinc finger protein we identified contains a N-terminal C2H2 zinc finger and an array of 7 zinc fingers in the C terminus. We used BindZFPredictor from DeepZF (42) to estimate the probability that each zinc finger is DNA binding or not. The first zinc finger had very low probability to bind DNA, which is not surprising as the N-terminals of proteins with C2H2 arrays often have protein-protein interaction domains in higher metazoans (39). The other 7 zinc fingers had a very high probability to bind DNA (Table 1). Another good predictor of DNA binding is the presence of TGEKP-like amino acid sequences between the zinc fingers (linkers) (43,44). Linkers act to place the zinc finger in the right position to bind DNA and mutations in the linkers have been associated with loss of affinity to DNA (44).

Table 1: C2H2 zinc fingers DNA binding probability generated using BindZFPredictor

Zinc Finger	Probability of binding DNA	Linker sequence
ZF1	14.0	- (VDGEP)
ZF2	84.3	TGEKP
ZF3	94.1	TGERP
ZF4	96.4	KGQKP
ZF5	98.1	TGQKP
ZF6	96.4	SGERT
ZF7	92.3	TGEKP
ZF8	65.5	- (QIKS-)

To explore whether a transcription factor binds the promoter of SUM and acts as an activator, or the promoter of Dsx and acts as a repressor in females, we searched for signatures of TF binding in both promoters and within the entire region using the published single nucleus ATACseq data from the A. franciscana female reproductive system. We used TOBIAS (45), which performs bias correction and detection of signatures of bound transcription factors (footprints) in chromatin accessibility data by looking for regions with coverage drops within regions of high coverage. The largest footprint in the Dsx surrounding region in A. franciscana was at the promoter of SUM, and no footprint was identified at the promoter of Dsx in both replicates of A. franciscana (Figure 5).

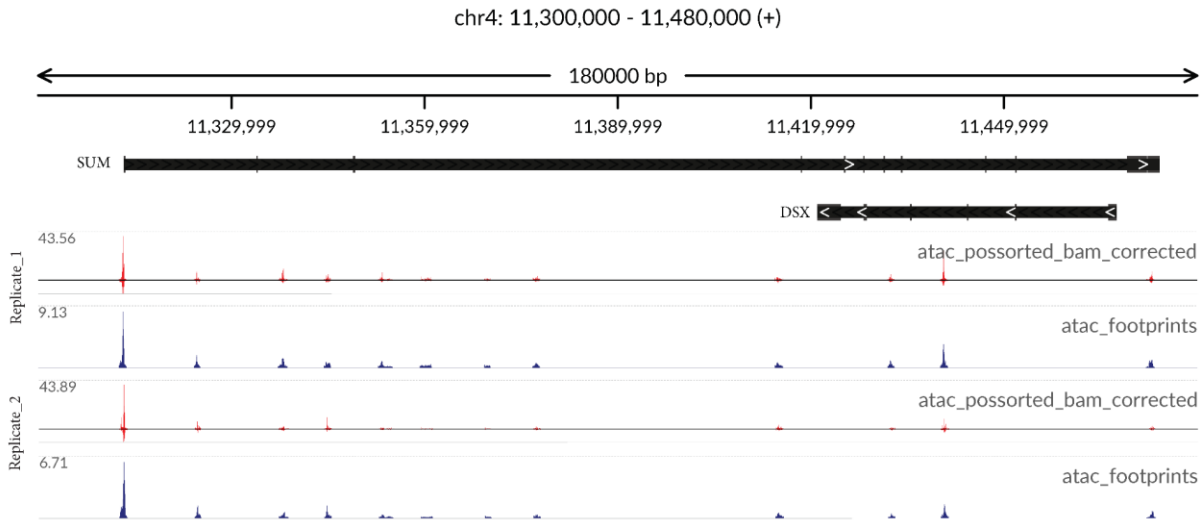


Figure 5: Footprinting using ATACseq data shows a large peak at the promoter of SUM.

The footprint suggests that a TF is bound in the promoter region of SUM. To check if FEM has a binding site there, we first predicted the binding sites for its 7 zinc fingers likely to bind DNA directly from the protein sequence using two different tools: the deepZF PWMpredictor and the Princeton Interactive PWM predictor (Supplementary Figures 5, 6 and 7). Both tools output a position weight matrix which can be used to score sequences for binding sites (Figure 5). As each zinc finger is known to bind 3 to 4 bp, we divided the promoter sequence into 31 bp k-mers, and we used PWM score from PWM tools to score each k-mer sequence (https://epd.expasy.org/cgi-bin/pwmtools/pwmscore). We also used the Alphafold3 server, which has shown promise in predicting protein-DNA interactions, to estimate the likelihood

each sequence (k-mer) would interact with FEM. We searched for regions which have a high footprint score, high predicted template modeling (pTM) and interface predicted template modeling (ipTM) scores, and high sequence conservation across the two species (A. fran and A. sinica). We identified a region, which seems to be relatively conserved across the two species, contains sequences that bind FEM in Silico with very confident predictions (AlphaFold) and also falls within the region with the high footprint score. The region also contains partially overlapping peaks of highly scoring sequences for the predicted PWMs (Figure 5). The PWM predicted by the two tools suggest that FEM binds A/T rich sequences; however, the scores from ZF princeton are more correlated with the AlphaFold ipTM values (pearson correlation=0.52, p-value=3.94e-15) than the predictions from DeepZF (pearson correlation=-0.026, p-value=0.72). All in all, this analysis suggests that FEM could be binding at the promoter of SUM and is therefore a promising candidate regulator of the SUM/Dsx locus.

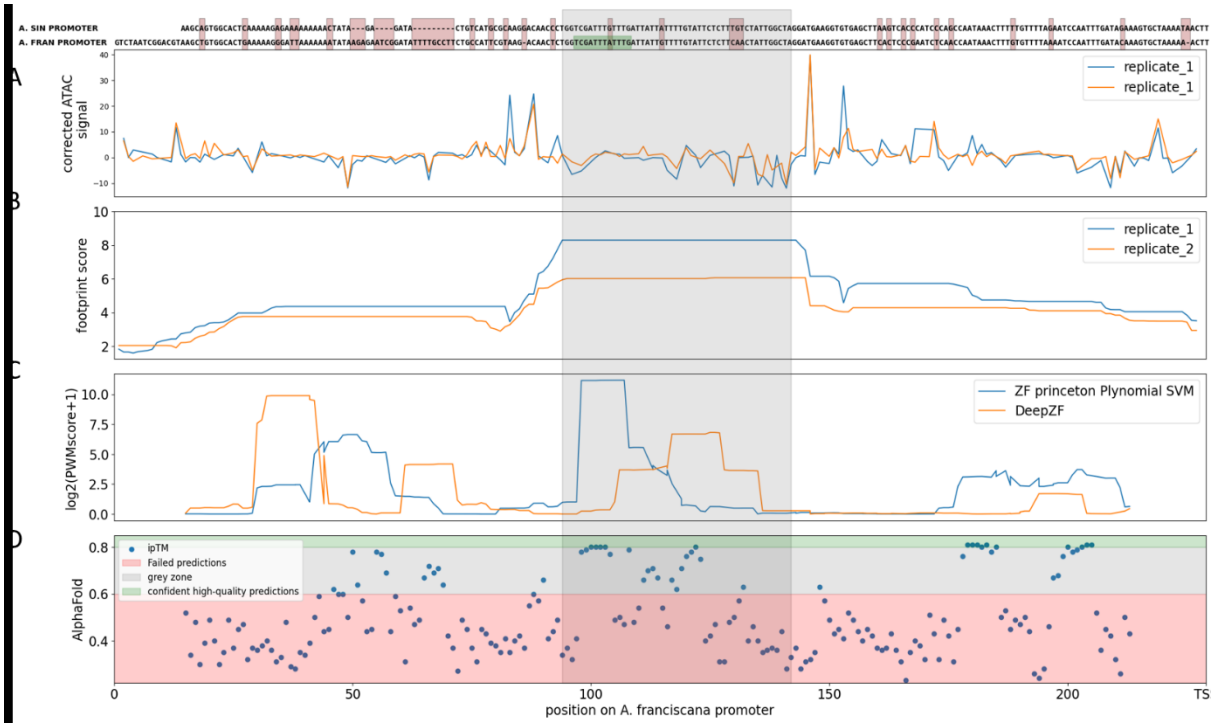


Figure 6: Binding analysis highlights a likely region for FEM binding in SUM promoter. A) promoter sequences of the two species (SNPs are highlighted, and - denote alignments gaps). B) corrected single nucleus ATACseq signal. C) Footprint score calculated using TOBIAS. C) PWM score for each K-mer estimated using the PWM generated using DeepZF and ZF princeton (the scores were shifted so that each point falls at the center of the K-mer) D) the predicted template modeling (pTM) and interface predicted template modeling (ipTM) scores for each K-mer and the 7 zinc finger array generated using AlphaFold.

In order to experimentally check if FEM is involved in the regulation of SUM and/or DSX, we attempted to use dsRNA to knockdown FEM in cells cultured from *A. parthenogenetica* and *Artemia franciscana* cysts (hydrated in 35 g/L salt water for 3 to 4 hours). To isolate the cells, we followed the protocol described in (46). Directly after isolation, we incubated the cells for 1 hour in L15 medium with 15 µg of dsRNA generated using the Jena HighYield T7 RNAi Kit (product size 348 bp) before adding FBS. We measured the viability and extracted RNA after two days of incubation from the two controls and two RNAi replicates from the two species, and we performed qPCR using multiple primers for FEM, DSX, and SUM. The preliminary results suggest that there is no decrease in the expression level of FEM, suggesting that any interference we observe likely occurs due to translational interference rather than RNA interference, as has been observed (47). Although the pattern for SUM is difficult to interpret, we see an increase in the expression of DSX in the cells of both species. This suggests that

the dsRNA may be having an effect on the cells, either by blocking the translation of FEM instead of degradation or through subtle downregulation of FEM that was undetected due to the amplification of the dsRNA and/or degradation products. Another possibility is that it takes more time to see a decrease in the expression of FEM and SUM compared to the slight increase we see in the expression of DSX, which should be completely absent in Aibi. This is ongoing work, and we plan to repeat the experiments with more replicates to check if the patterns hold.

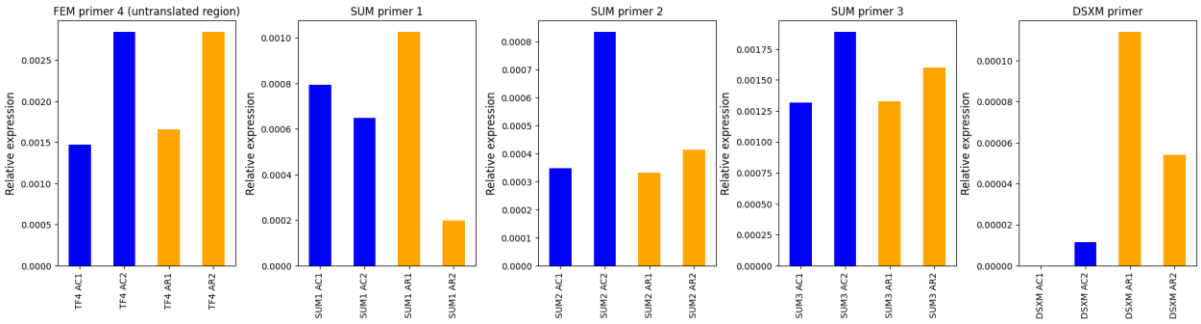


Figure 7: A) RNAi analysis in A. parthenogenetica cells suggests FEM dsRNA might be negatively regulating the expression of Dsx.

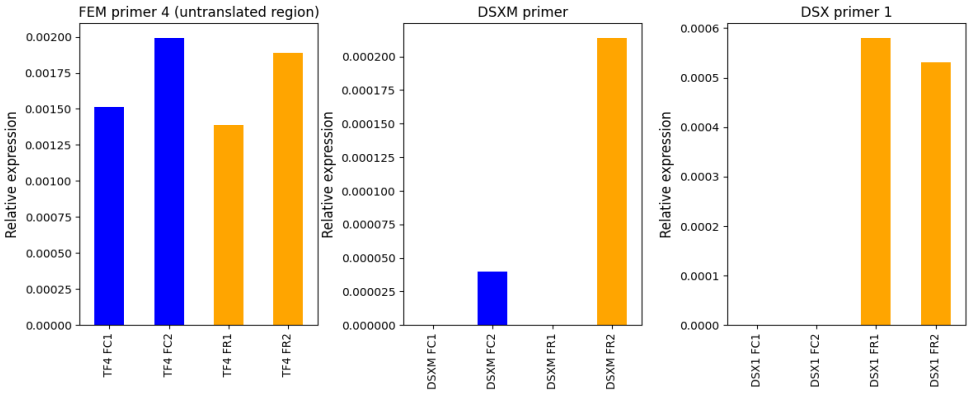


Figure 7: B) RNAi analysis in A. franciscana cells suggests FEM dsRNA might be negatively regulating the expression of Dsx.

Discussion

The role of long noncoding RNAs and overlapping genes in gene regulation

Long noncoding RNA, often defined as transcripts longer than 200 bp, perform various complex biological roles including regulating nearby genes by recruiting regulatory factors and/or interfering with the RNA pol II recruitment to the promoter of an overlapping gene (cis-acting) (48,49). They can also regulate distant genes by associating with other proteins like chromatin-modifying complexes (trans-acting) (48). They evolve more rapidly than protein coding genes, which makes identifying and comparing them across species very difficult (30). Many examples exist of long noncoding RNA with important sex-specific roles: the roX1 and roX2 RNAs, for instance, have been shown to be essential components of the dosage compensation complex in *Drosophila* (50,51). In mammals, the long noncoding RNA Xist mediates dosage compensation by accumulating on one of the X chromosomes and triggering a cascade of events that leads to its silencing (52). In this paper, we have identified a long noncoding RNA, which overlaps and potentially regulates the expression of Doublesex, which is known to be essential for male differentiation across all Arthropods. In *Artemia franciscana*, knocking-down Dsx using RNAi has been shown to cause males to produce ovary-like structures (20). The sequence-level analysis suggests that various exons of the antisense gene SUM are not conserved, which argues against a conserved sequence-dependent

function for the lncRNA. The presence of various histone modifications at the promoter of SUM, on the other hand, could suggest that its recruitment of chromatin-modifying factors (Figure 4) and consequent transcription might be directly interfering with the expression of Dsx. This is further supported by the fact that in the *Artemia* paper, interfering with SUM, which was previously suspected to be a female isoform of Dsx produced by alternative splicing, had no phenotypic effect (20). However, this does not fully rule out a female specific function for the various produced transcripts from the locus. One of the most similar examples to what we describe is the SUPPRESSOR OF FEMINIZATION (SUF) long non coding RNA identified in the Liverwort Marchantia polymorpha. SUF overlaps the female promoting gene FGMYB and suppresses its expression by its mere transcription, leading to male development (49). Although we suspect that SUM acts as a suppressor in a similar manner by the transcription of the overlapping region, once gene editing becomes possible in *Artemia*, it would be possible to directly test this by the insertion of a transcriptional terminator that would stop the transcription into the DSX region, as was done in Marchantia.

C2H2 zinc fingers complex roles and diverse functions in gene regulation

Protein domains coordinated by a zinc atom are mainly found in eukaryotes, and within the more than 20 different domains identified, C2H2 zinc fingers are the most abundant (39). The zinc finger families presumably expanded and diversified through gene duplication, mutation, and selection (53). They are primarily known for binding DNA, but many examples exist of C2H2 domains interacting with RNA and protein (54,55). In Humans, more than 800 zinc finger containing proteins have been identified (with a median number of 6 C2H2 domains per protein) and in *Drosophila* more than 300 C2H2 proteins have been identified (with a median number of 6 C2H2 domains per protein) (56,57). The properties of zinc finger transcription factors and their ability to recognize specific sequences makes them great candidates for mediating sex-specific regulatory functions. A genome-wide study of C2H2 zinc finger proteins in Bombyx mori has identified several with sex-specific functions in metamorphosis and female reproduction (58). Another well-studied example of a C2H2 zinc finger protein is the Chromatin-linked adaptor for MSL proteins (CLAMP) in *Drosophila*, which associates with the other components of the dosage compensation complex through its N-terminal domain and is involved in the recruitments of the complex to the specific sites on the X-chromosome using the array of zinc fingers in its C-terminal domain (54,39). It is also possible that FEM associates with one or more proteins and binds cooperatively, as in the case of the MSL2 and CLAMP. It is tempting to draw parallels between CLAMP and FEM, partly due to the similarity in structure and partly due to the fact that in the two species the acetylation of H4K16 seems to have been convergently recruited for dosage compensation. The presence of a differentially accessible peak for H4K16ac at the promoter of SUM points in that direction as well (Figure 4). It is also possible that FEM associates with different complexes and performs various functions in females; maternally deposited CLAMP, for instance, has been shown to be involved in zygotic genome activation with Zelda (59).

DSX, a divergently regulated conserved regulator

The DMRT (doublesex- and mab-3-related transcription factor) family of genes are characterized by the presence of a DM DNA binding domain (an intertwined zinc finger-like DNA binding module), and they are found in all vertebrate and invertebrate genomes studied (60). The domain acquired its name after being discovered in Doublesex in *Drosophila* and male abnormal-3 (mab-3) in *C. elegans*, with a common evolutionary origin. The mutant phenotypes that led to their discoveries suggested an important role sex-specific differentiation, and since then, the DMRT family genes have been found to be involved in sexual differentiation across all studied metazoans (61). DM genes are also associated with sex reversal in humans (62). Binding analyses of DMRT binding domains show that they have similar binding preferences across distant species (63). What is as puzzling as the conservation of the role of DMRT genes at the interface between sex determination and differentiation (60) is the extreme diversity of mechanisms that evolved to regulate their expression and functions. In arthropods, the regulation of Doublesex by alternative splicing is

the best characterized case, but recent studies have shown that even within insects transitions to male-specific transcription have happened, as in the case of termites (61). Moreover, many novel regulatory mechanisms have been discovered in non-insect arthropods, such as *Daphnia*, where Vrille, an ortholog of the circadian clock genes, has been co-opted to regulate Dsx (64) along with an overlapping long noncoding RNA (65). A recent analysis in mud crab suggested that miR-34, a microRNA, negatively regulates Dsx and other Dmrt genes in the testis (66). In this study, we report a novel mechanism for the regulation of Dsx through the transcription of an overlapping non-coding RNA, adding to the diversity of mechanisms reported. It is not clear at this stage whether this mechanism evolved in *Artemia* or is shared by other Anostracans, and a detailed analysis of regulatory mechanisms in other brachiopods will be needed to shed light on its evolutionary dynamics in this group.

Materials and Methods

Sampling, RNA extractions and sequencing

We isolated pairs of males and females of *Artemia sinica* and *Artemia franciscana* from the colonies maintained in our lab in individual vials and checked the vials for nauplii at 24 hour intervals. We froze 14 nauplii from one pair of *Artemia sinica* and 14 nauplii from one pair of *Artemia franciscana*. The RNA was extracted using Maxwell RSC simplyRNA tissue kit and sent to the Vienna BioCenter for library preparation and sequencing. The *A. parthenogenetica* rare male samples were processed the same way.

Identification of the W-specific transcripts in A. sinica

We used BBMap bbduk.sh (67) to identify female specific RNA kmers in nauplii and adults separately. The paired RNA reads enriched for the female specific kmers were then recovered and assembled using Trinity. We used the trinity script (Trinity_gene_splice_modeler.py) to merge the isoforms and generate supergenes. We then mapped all the male and female DNA kmers to the supergenes and filtered them using the number of perfectly matching male and female kmers. We merged the W-candidates resulting from the nauplii and adult analyses and removed the duplicates.

Genome annotation and differential expression analysis:

For *Artemia franciscana*, we mapped the RNAseq reads using Hisat2 (with –fr for nauplii and –rf for adults for –strandedness). We used the resulting bam files to generate GTFs and merge them using StringTie2 (TPM >0.5). We used the lncDC tool (trained using the *Drosophila* coding and long noncoding RNA) to identify mRNA and lncRNAs and add the biotype per transcript to the GTF. For *Artemia sinica*, we added the putative W supergenes to the genome and annotated it using the same approach as A. franciscana. We used StringTie2 in the quantification mode (-e) to get the counts for each gene using the merged GTFs, and we used DEseq2 to perform the differential expression analysis. We extracted the longest isoform for each gene and used the blat to map the transcriptomes of the two species. We identified the best match using the 2-besthitblat.pl script and the reciprocal best hits using reciprocal_besthit_python.py.

PCR and RT-qPCR

For all PCRs and qPCRs performed in this study, we used the Maxwell RSC simplyRNA tissue kit for RNA extractions, and then we used the High-Capacity RNA-to-cDNA Kit (Thermofisher 4388950) to prepare the cDNA. The PCRs were performed using the NEB kit (Phusion Hot Start Flex 2X Master Mix), and the qPCRs using the promega (GoTaq qPCR Master Mix).

ATACseq footprinting

We treated the single nucleus ATACseq data of the *Artemia franciscana* female reproductive system as bulk ATACseq and used the TOBIAS ATACorrect command with the bam file and ATACseq peaks generated by CellRanger. The generated Bigwig files were then used with

the FootprintScores commands to identify the regions likely to be bound by transcription factors.

Zinc Finger Binding Analysis

We identified the zinc fingers using both PROSITE and NCBI CDD. We then generated the csv file with each zinc finger and the 40 adjacent amino acids appended to each side (or the largest number available) and used the DeepZF BindZFPredictor to estimate the probability of each zinc finger binding DNA. We used the PWM predictor to generate the PWM for the zinc fingers with >0.5 probability of binding DNA using both the DeepZF PWMpredictor and ZF princeton (https://zf.princeton.edu/).

To look for binding sequences in the promoter of SUM, we extracted the 229 bp upstream of the putative transcription start site and used the kmercountexact.sh script from bbmap to generate all the 31 bp k-mers. We score all the kmers against the generated PWMs using PWM score (https://epd.expasy.org/pwmtools/pwmscore.php). Additionally, we extracted the region with the 7 zinc finger from the protein sequences and used RoseTTAFoldNA and the AlphaFold server to predict the interaction with each K-mer.

dsRNA synthesis, *Artemia* cell culture, and RNAi

We synthesized dsRNA for RNAi using Jena HighYield T7 RNAi Kit. The DNA template was prepared with PCR using FEM primers containing opposite T7 promoters at the 5' ends of each strand. The yield was quantified using Nanodrop. For the cell culture, we hydrated the cysts using 35 g/L red sea salt for 3-4 hours, and then we removed the water and added 2% NaClO (clorox). When the cysts turned orange, 8 mL of Na2S2O3 (1% final concentration) was added to stop the reaction. We rinsed the cysts using 15 ml 3% seawater and then 15 ml of 3X PBS. The PBS was removed and the cysts were moved to an eppendorf, where they were gently ground in L-15 medium using a pestle. Once the mixture became cloudy, 800 ul of L-15 medium were added and the mixture was centrifuged for 10 minutes (1000 g at 4 °C). The supernatant was then filtered using a 35 µm FACs strainer and moved to the well of a 24-well plate. We then added 637.5uL of L-15 medium to each well and 15 ug of dsRNA (5 uL of 3000 ng/uL), and incubated for an hour before adding 112.5uL of FBS.

References

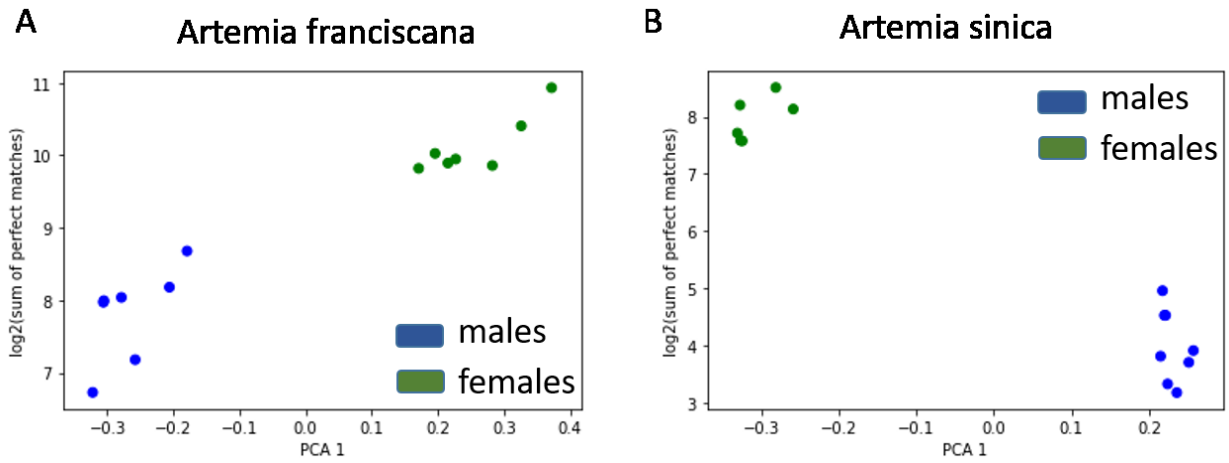
1. Pan Q, Kay T, Depincé A, Adolfi M, Schartl M, Guiguen Y, et al. Evolution of master sex determiners: TGF-β signalling pathways at regulatory crossroads. Philos Trans R Soc B Biol Sci. 2021 Jul 12;376(1832):20200091.
2. Yu H, Du X, Chen X, Liu L, Wang X. Transforming growth factor-β (TGF-β): A master signal pathway in teleost sex determination. Gen Comp Endocrinol. 2024 Sep 1;355:114561.
3. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, et al. Sex Determination: Why So Many Ways of Doing It? PLoS Biol. 2014 Jul 1;12(7):e1001899.
4. Adolfi MC, Herpin A, Schartl M. The replaceable master of sex determination: bottom-up hypothesis revisited. Philos Trans R Soc B Biol Sci. 376(1832):20200090.
5. Garcia-Alonso L, Lorenzi V, Mazzeo CI, Alves-Lopes JP, Roberts K, Sancho-Serra C, et al. Single-cell roadmap of human gonadal development. Nature. 2022 Jul;607(7919):540–7.
6. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical Compensation of Zygotic X Transcription in Early *Drosophila melanogaster* Development Revealed through Single-Embryo RNA-Seq. PLOS Biol. 2011 Feb 8;9(2):e1000590.
7. Kopp A. Dmrt genes in the development and evolution of sexual dimorphism. Trends Genet. 2012 Apr;28(4):175–84.
8. Baral S, Arumugam G, Deshmukh R, Kunte K. Genetic architecture and sex-specific selection govern modular, male-biased evolution of doublesex. Sci Adv. 2019 May 8;5(5):eaau3753.
9. Chikami Y, Okuno M, Toyoda A, Itoh T, Niimi T. Evolutionary History of Sexual Differentiation Mechanism in Insects. Mol Biol Evol. 2022 Jul 1;39(7):msac145.
10. Han C, Peng Q, Su X, Xing L, Ji X, Pan Y. A male-specific doublesex isoform reveals an evolutionary pathway of sexual development via distinct alternative splicing mechanisms. Commun Biol. 2022 Jul 22;5(1):1–11.
11. Zhang P, Yang Y, Xu Y, Cui Z. Analyses of the Dmrt family in a decapod crab, Eriocheir sinensis uncover new facets on the evolution of DM domain genes. Front Physiol. 2023 May 26;14:1201846.
12. Ye Z, Bishop T, Wang Y, Shahriari R, Lynch M. Evolution of sex determination in crustaceans. Mar Life Sci Technol. 2023 Feb 1;5(1):1–11.
13. Bowen ST. THE GENETICS OF ARTEMIA SALINA. III. EFFECTS OF X-IRRADIATION AND OF FREEZING UPON CYSTS. Biol Bull. 1963 Dec;125(3):431–40.
14. 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. Liu Z, editor. PLoS ONE. 2013 Mar 4;8(3):e57585.
15. Huylmans AK, Toups MA, Macon A, Gammerdinger WJ, Vicoso B. Sex-Biased Gene Expression and Dosage Compensation on the *Artemia franciscana* Z-Chromosome. Mank J, editor. Genome Biol Evol. 2019 Apr 1;11(4):1033–44.
16. Elkrewi M, Khauratovich U, Toups MA, Bett VK, Mrnjavac A, Macon A, et al. ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp. Dyer K, editor. Genetics. 2022 Sep 30;222(2):iyac123.
17. Bett VK, Macon A, Vicoso B, Elkrewi M. Chromosome-Level Assembly of *Artemia franciscana* Sheds Light on Sex Chromosome Differentiation. Genome Biol Evol [Internet]. 2024 Jan 5 [cited 2024 Feb 5];16(1). Available from: https://dx.doi.org/10.1093/gbe/evae006
18. Jo E, Lee SJ, Choi E, Kim J, Lee JH, Park H. Sex-Biased Gene Expression and Isoform Profile of Brine Shrimp *Artemia franciscana* by Transcriptome Analysis. Anim Open Access J MDPI. 2021 Sep 7;11(9):2630.
19. Viet DN, Christiaens O, De Vos S, Smagghe G, Bossier P. The Sex-Specific Splicing of Doublesex in Brine Shrimp *Artemia franciscana*. Genes. 2022 Nov 1;13(11):1997.
20. Wu WT, Xu LY, Yan ZJ, Bi N, Cheng CY, Yang F, et al. Identification and characterization of the Doublesex gene and its mRNA isoforms in the brine shrimp *Artemia franciscana*. Biochem J. 2023 Mar 15;480(5):385–401.

21. Kalita AI, Marois E, Kozielska M, Weissing FJ, Jaouen E, Möckel MM, et al. The sex-specific factor SOA controls dosage compensation in Anopheles mosquitoes. Nature. 2023 Nov;623(7985):175–82.
22. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug;37(8):907–15.
23. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019 Dec 16;20(1):278.
24. 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. Wilson M, editor. Mol Biol Evol. 2021 Dec 9;38(12):5345–58.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550.
26. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.
27. Elkrewi M, Vicoso B. Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome. PLOS Genet. 2024 Aug 30;20(8):e1011376.
28. Huylmans AK, Macon A, Hontoria F, Vicoso B. Transitions to asexuality and evolution of gene expression in *Artemia* brine shrimp. Proc Biol Sci. 2021 Sep 29;288(1959):20211720.
29. Li M, Liang C. LncDC: a machine learning-based tool for long non-coding RNA detection from RNA-Seq data. Sci Rep. 2022 Nov 9;12(1):19083.
30. Szcześniak MW, Kubiak MR, Wanowska E, Makałowska I. Comparative genomics in the search for conserved long noncoding RNAs. Essays Biochem. 2021 Oct;65(4):741–9.
31. Mohammadin S, Edger PP, Pires JC, Schranz ME. Positionally-conserved but sequence-diverged: identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. BMC Plant Biol. 2015 Sep 11;15(1):217.
32. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Apr 29;10(1):1930.
33. Yashar WM, Kong G, VanCampen J, Curtiss BM, Coleman DJ, Carbone L, et al. GoPeaks: histone modification peak calling for CUT&Tag. Genome Biol. 2022 Jul 4;23(1):144.
34. Benveniste D, Sonntag HJ, Sanguinetti G, Sproul D. Transcription factor binding predicts histone modifications in human cell lines. Proc Natl Acad Sci. 2014 Sep 16;111(37):13367–72.
35. Patiyal S, Tiwari P, Ghai M, Dhapola A, Dhall A, Raghava GPS. A hybrid approach for predicting transcription factors. Front Bioinforma. 2024;4:1425419.
36. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. Nucleic Acids Res. 2013 Jan 1;41(D1):D344–7.
37. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015 Jan 28;43(D1):D222–6.
38. Zhang X, Blumenthal RM, Cheng X. Updated understanding of the protein–DNA recognition code used by C2H2 zinc finger proteins. Curr Opin Struct Biol. 2024 Aug 1;87:102836.
39. Bonchuk AN, Georgiev PG. C2H2 proteins: Evolutionary aspects of domain architecture and diversification. BioEssays News Rev Mol Cell Dev Biol. 2024 Aug;46(8):e2400052.
40. Jen J, Wang YC. Zinc finger proteins in cancer progression. J Biomed Sci. 2016 Jul 13;23(1):53.
41. Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, et al. Zinc-finger proteins in health and disease. Cell Death Discov. 2017 Nov 13;3(1):1–12.
42. Aizenshtein-Gazit S, Orenstein Y. DeepZF: improved DNA-binding prediction of C2H2-zinc-finger proteins by deep transfer learning. Bioinformatics. 2022 Sep 16;38(Supplement_2):ii62–7.
43. Brayer KJ, Segal DJ. Keep Your Fingers Off My DNA: Protein–Protein Interactions

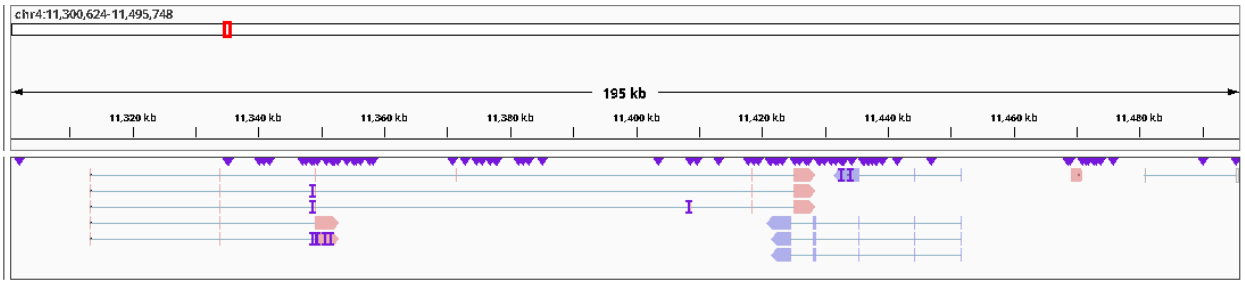
Mediated by C2H2 Zinc Finger Domains. Cell Biochem Biophys. 2008 Mar 1;50(3):111–31.
44. Hamed MY, Siam R, Zaid R. The role of zinc finger linkers in zinc finger protein binding to DNA. J Comput Aided Mol Des. 2021 Sep;35(9):973–86.
45. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation | Nature Communications [Internet]. [cited 2024 Dec 31]. Available from: https://www.nature.com/articles/s41467-020-18035-1
46. Duan H, Jin Y, Shao X, Sun P, Wang X, Sui L. Stable primary embryonic cells of *Artemia* are suitable for tracing the process of *V. anguillarum* and *V. parahaemolyticus* infection. Aquaculture. 2022 Nov 15;560:738598.
47. López-Fraga M, Martínez T, Jiménez A. RNA interference technologies and therapeutics: from basic research to products. BioDrugs Clin Immunother Biopharm Gene Ther. 2009;23(5):305–32.
48. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021 Feb;22(2):96–118.
49. Kajiwara T, Miyazaki M, Yamaoka S, Yoshitake Y, Yasui Y, Nishihama R, et al. Transcription of the Antisense Long Non-Coding RNA, SUPPRESSOR OF FEMINIZATION, Represses Expression of the Female-Promoting Gene FEMALE GAMETOPHYTE MYB in the Liverwort Marchantia polymorpha. Plant Cell Physiol. 2024 Mar 1;65(3):338–49.
50. Franke A, Baker BS. The *rox1* and *rox2* RNAs Are Essential Components of the Compensasome, which Mediates Dosage Compensation in *Drosophila*. Mol Cell. 1999 Jul 1;4(1):117–22.
51. Kim M, Faucillion ML, Larsson J. RNA-on-X 1 and 2 in *Drosophila melanogaster* fulfill separate functions in dosage compensation. PLOS Genet. 2018 Dec 10;14(12):e1007842.
52. Cerase A, Pintacuda G, Tattermusch A, Avner P. Xist localization and function: new insights from multiple levels. Genome Biol. 2015 Aug 15;16(1):166.
53. Emerson RO, Thomas JH. Adaptive Evolution in Zinc Finger Transcription Factors. PLoS Genet. 2009 Jan 2;5(1):e1000325.
54. Tikhonova E, Mariasina S, Efimov S, Polshakov V, Maksimenko O, Georgiev P, et al. Structural basis for interaction between CLAMP and MSL2 proteins involved in the specific recruitment of the dosage compensation complex in *Drosophila*. Nucleic Acids Res. 2022 Jun 24;50(11):6521–31.
55. Nabeel-Shah S, Pu S, Burns JD, Braunschweig U, Ahmed N, Burke GL, et al. C2H2-zinc-finger transcription factors bind RNA and function in diverse post-transcriptional regulatory processes. Mol Cell. 2024 Oct 3;84(19):3810-3825.e10.
56. Jauch R, Bourenkov GP, Chung HR, Urlaub H, Reidt U, Jäckle H, et al. The Zinc Finger-Associated Domain of the *Drosophila* Transcription Factor Grauzone Is a Novel Zinc-Coordinating Protein-Protein Interaction Module. Structure. 2003 Nov 1;11(11):1393–402.
57. Zuo Z, Billings T, Walker M, Petkov PM, Fordyce PM, Stormo GD. On the dependent recognition of some long zinc finger proteins. Nucleic Acids Res. 2023 Jun 23;51(11):5364–76.
58. Wu S, Tong X, Li C, Lu K, Tan D, Hu H, et al. Genome-wide identification and expression profiling of the C2H2-type zinc finger protein genes in the silkworm Bombyx mori. PeerJ. 2019 Jul 5;7:e7222.
59. Duan J, Rieder L, Colonnetta MM, Huang A, Mckenney M, Watters S, et al. CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. Yamashita YM, Struhl K, editors. eLife. 2021 Aug 3;10:e69937.
60. Verhulst EC, van de Zande L. Double nexus—Doublesex is the connecting element in sex determination. Brief Funct Genomics. 2015 Nov;14(6):396–406.
61. Miyazaki S, Fujiwara K, Kai K, Masuoka Y, Gotoh H, Niimi T, et al. Evolutionary transition of doublesex regulation from sex-specific splicing to male-specific transcription in termites. Sci Rep. 2021 Aug 6;11(1):15992.
62. Zhu L, Wilken J, Phillips NB, Narendra U, Chan G, Stratton SM, et al. Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. Genes Dev. 2000 Jul 15;14(14):1750–64.
63. Murphy MW, Zarkower D, Bardwell VJ. Vertebrate DM domain proteins bind similar

DNA sequences and can heterodimerize on DNA. BMC Mol Biol. 2007 Jul 2;8:58.
64. Ishak NSM, Nong QD, Matsuura T, Kato Y, Watanabe H. Co-option of the bZIP transcription factor Vrille as the activator of Doublesex1 in environmental sex determination of the crustacean *Daphnia magna*. PLOS Genet. 2017 Nov 2;13(11):e1006953.
65. Kato Y, Perez CAG, Ishak NSM, Nong QD, Sudo Y, Matsuura T, et al. A 5′ UTR-Overlapping LncRNA Activates the Male-Determining Gene doublesex1 in the Crustacean *Daphnia magna*. Curr Biol. 2018 Jun 4;28(11):1811-1817.e4.
66. Huang J, Wan H, Jiang J, Huang Y, Zou P, Zhang Z, et al. miR-34 negatively regulates the expression of *Dmrt* and related genes in the testis of mud crab *Scylla paramamosain*. Comp Biochem Physiol B Biochem Mol Biol. 2025 Jan 1;275:111018.
67. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner [Internet]. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States); 2014 Mar [cited 2023 Sep 4]. Report No.: LBNL-7065E. Available from: https://www.osti.gov/biblio/1241166

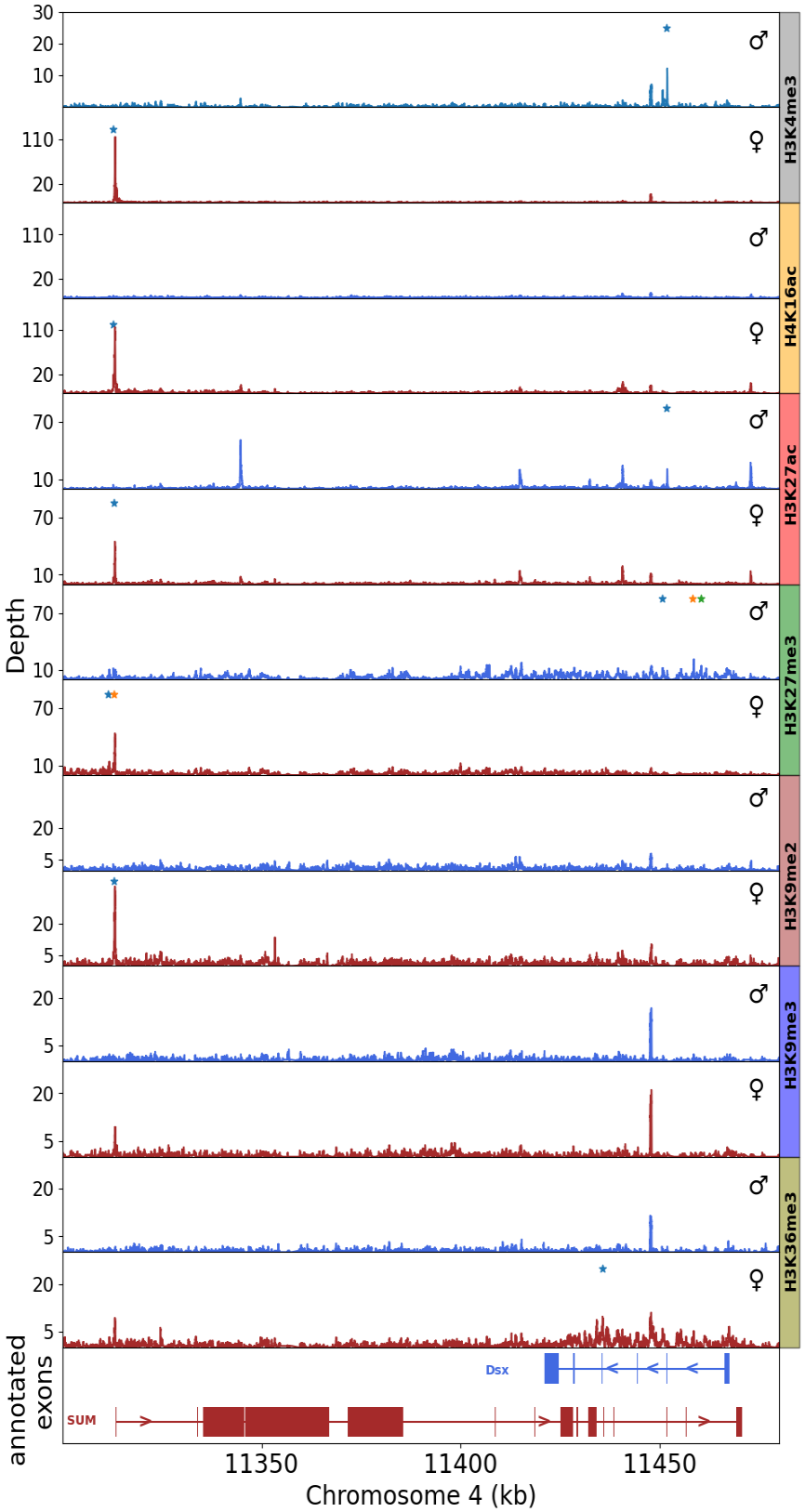
Supplementary material



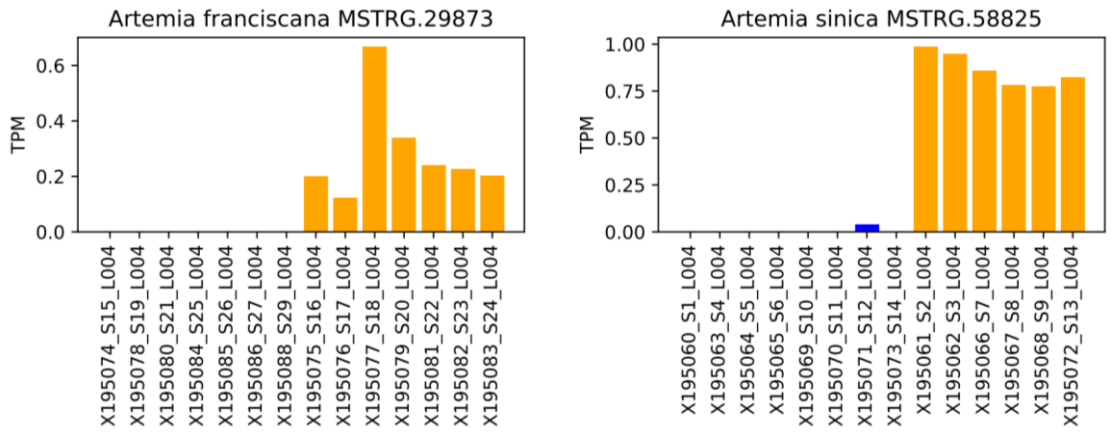
Supplementary figure 1: PCA and perfect matches analysis to identify the sex of the nauplii. A) Artemia franciscana B) Artemia sinica



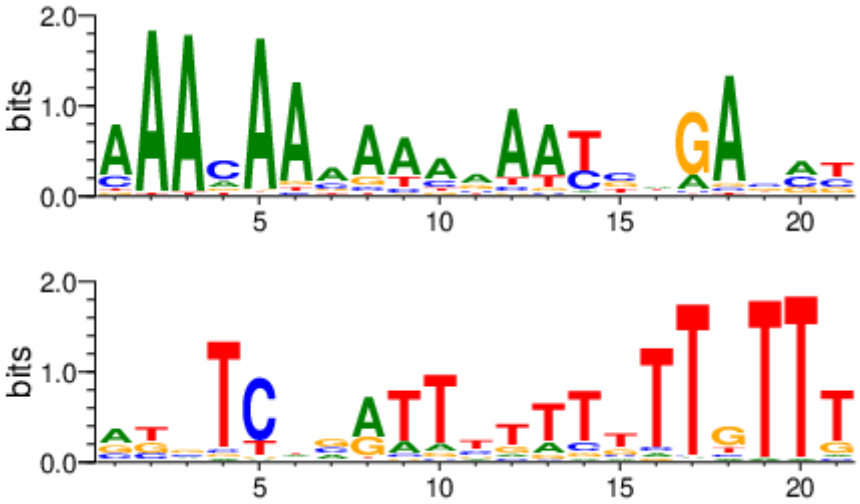
Supplementary figure 2: Integrative genome view image showing the long reads mapping in the Dsx region and confirming the presence of the convergently overlapping region.



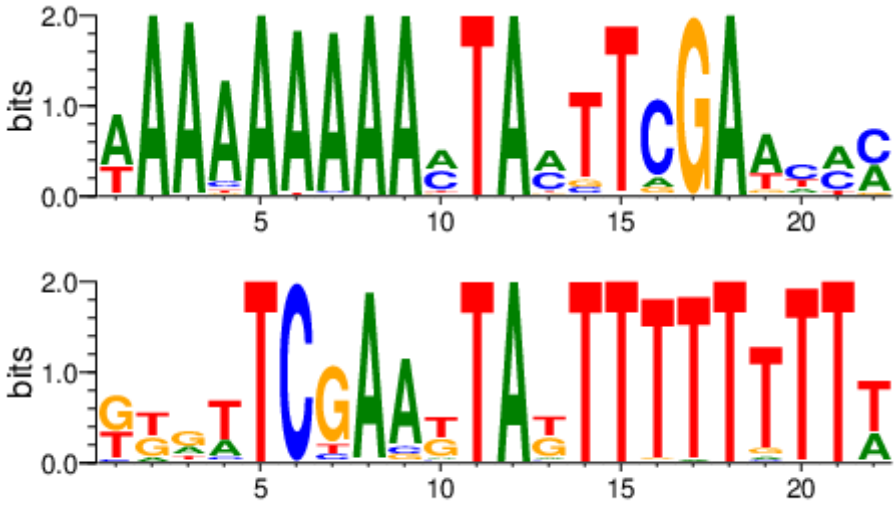
Supplementary figure 3: histone modifications in the DSX and SUM region.



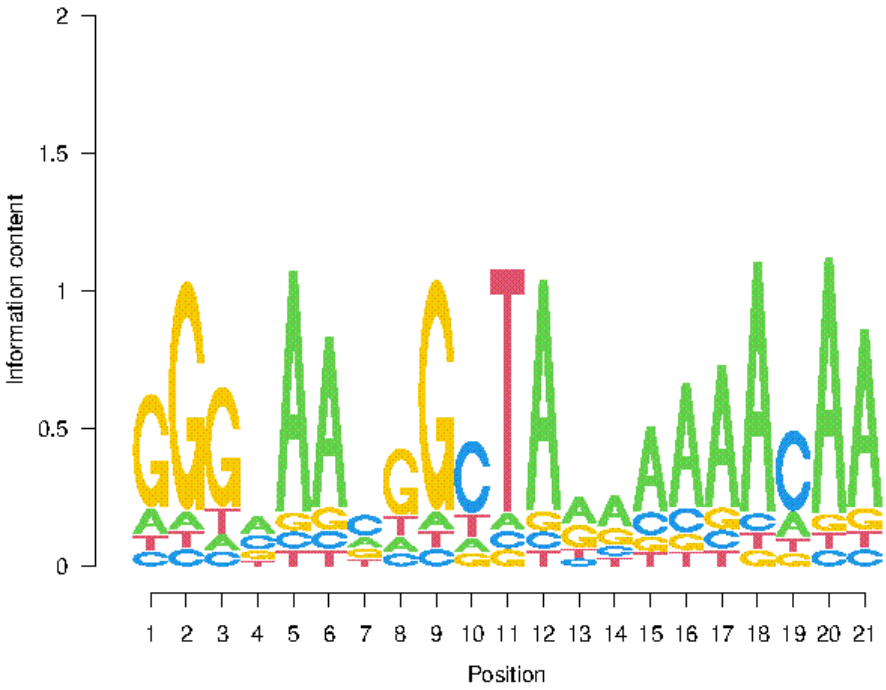
Supplementary figure 4: The expression of FEM in the nauplii samples.



Supplementary Figure 5: ZF princeton predicted position weight matrix (PWM using RF regression on B1H)



Supplementary Figure 6: ZF princeton predicted position weight matrix (PWM using polynomial SVM)



Supplementary Figure 7: DeepZF PWM predicted position weight matrix (PWM).

CHAPTER 4

Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome

Published

Chapter 4: Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome

Authors:
Marwan Elkrewi[1, *], Beatriz Vicoso[1, *]
* marwanelkrewi@gmail.com (ME), bvicoso@ist.ac.at (BV)
[1] Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria

Key words:
Artemia single-nucleus Atlas, Meiosis, Germline development, Dosage compensation, MSCI

Abstract

Our understanding of the molecular pathways that regulate oogenesis and define cellular identity in the Arthropod female reproductive system and the extent of their conservation is currently very limited. This is due to the focus on model systems, including *Drosophila* and *Daphnia*, which do not reflect the observed diversity of morphologies, reproductive modes, and sex chromosome systems. We use single-nucleus RNA and ATAC sequencing to produce a comprehensive single nucleus atlas of the adult *Artemia franciscana* female reproductive system. We map our data to the Fly Cell Atlas single-nucleus dataset of the *Drosophila melanogaster* ovary, shedding light on the conserved regulatory programs between the two distantly related Arthropod species. We identify the major cell types known to be present in the *Artemia* ovary, including germ cells, follicle cells, and ovarian muscle cells. Additionally, we use the germ cells to explore gene regulation and expression of the Z chromosome during meiosis, highlighting its unique regulatory dynamics and allowing us to explore the presence of meiotic sex chromosome silencing in this group.

Author summary

Oogenesis is a highly complex process involving multiple cell-types and an extremely well orchestrated program that unfolds in the female reproductive system. Despite the large diversity of Arthropod reproductive modes and sex determination systems, our current understanding of oogenesis is limited to a few model species. This makes it difficult to study and formulate hypotheses about the evolutionary history, constraints, and importance of the individual elements of this process. To fill this gap, we used single-nucleus expression and chromatin-accessibility data to produce a single-nucleus atlas of the *Artemia franciscana* female reproductive system. By comparing our dataset to the published *Drosophila* single-nucleus data (over 400 million years of divergence), we were able to highlight the substantial conservation of several of the molecular pathways of oogenesis and meiosis. We found evidence of global transcriptional quiescence and chromatin condensation in late germ cells, highlighting the conserved role of this repressive stage in arthropod oogenesis. Additionally, we explored the expression patterns of the ZW sex chromosomes during oogenesis. Our data shows that the Z-chromosome is consistently downregulated in germline cells. While this is partly driven by a lack of dosage compensation in the germline, a subset of cells show stronger repression of the Z chromosome.

Introduction

Most animals have evolved sexually dimorphic mechanisms and tissues dedicated to the production of haploid gametes through meiosis (gametogenesis). Males produce motile nuclei (sperm) through spermatogenesis, which takes place in the testes. Females produce oocytes that contain a haploid nucleus along with the cytoplasmic molecules needed to initiate and facilitate embryonic development through oogenesis, which takes place in the ovaries (1). While many aspects of oogenesis and female meiosis are highly conserved between distant species, there is considerable diversity in many others, including the presence/absence of nurse cells, meiotic chromosome pairing strategies, recombination rates, timing and duration of meiotic arrests, and sex chromosome specific regulation (1–3). Why such a fundamental and ancient mechanism exhibits so much variation is still unclear. Studying species with diverse body plans, reproductive modes, and sex chromosome systems at the genetic and molecular levels will help elucidate the developmental constraints and selective pressures that shaped the evolution of the conserved, convergent and divergent features of oogenesis.

As the hallmark of oogenesis, meiosis is very tightly regulated and many mechanisms have evolved to ensure the faithful transmission of genetic information to the offspring through the proper pairing and segregation of homologous chromosomes (4). In addition to ensuring the fidelity of the transmitted genome copy and building the maternal reserves to kick-start the embryo's journey, oogenesis involves extensive reprogramming of the epigenetic landscape to ensure a successful oocyte-to-embryo transition upon fertilization (5). To accomplish those feats, oocytes actively navigate previtellogenesis and the majority of prophase I before they arrest and become transcriptionally quiescent. The prophase I arrest is thought to be essential for oocyte growth and differentiation, and is conserved across metazoans, but with highly variable durations (6). In many arthropods, the lack of transcription in the oocyte is buffered by the activity of nurse cells that remain connected to the oocyte through cytoplasmic bridges (7). Sister cyst cells have been shown to play a similar role in mice, suggesting they might also be important for mammalian oocyte differentiation (8). In addition to this buffering, work in *Drosophila* suggests that some oocyte specific transcription, regulated through epigenetic programming in early oogenesis, occurs before the resumption of meiosis (9). This, along with the findings that oocyte chromatin enrichments of H4K16ac and H3K27me3 are maintained in the oocyte-to-embryo transitions in *Drosophila* and mammals, highlights the importance of the epigenetic regulation in early oogenesis (10–12). Somatic-germ cell signaling is also known to play an important role in oogenesis, where in *Drosophila* and mammals, signaling, often hormonal, by the surrounding follicle cells plays an important role in triggering oocyte maturation (13,14). Despite our understanding of some of the pathways involved, many questions still surround the intrinsic and extrinsic signaling involved in oocyte differentiation, maturation and the resumption of meiosis after the long arrest, including their relative contribution, how they evolved in the first place, and how conserved they are across metazoans.

The presence of differentiated/heteromorphic sex chromosomes, such as the X and Y pair of mammals, introduces two challenges: First, X-linked genes are found in different copy numbers between males and females, which can cause imbalances in expression; second, sequence similarity is usually required for successful synapsis and accurate segregation of homologous chromosomes, but is lacking over most of the length of the XY pair. Dosage compensation mechanisms have evolved to tackle the first problem, with some species, such as *Drosophila*, upregulating the expression of the single X in males, and others, such as mammals, inactivating one of the X chromosomes in females (15). This sex chromosome-specific epigenetic regulation seems to disappear/reverse in the germline cells of *Drosophila* males (16,17) and undergo extensive reconfiguration in mammalian females, with a period of hyper-transcription before reaching dosage balance (X:A ~ 1) (18,19). It is not clear whether this absence of dosage compensation in the germline is linked to the epigenetic reprogramming that takes place during gametogenesis, and whether it is the rule or the exception in Arthropods.

The second problem is germline specific, as the sex chromosomes fail to synapse across some or most of their length during prophase I (20). Asynapsis of other chromosomes triggers germ cell arrest as a defense mechanism against genome instability and aneuploidy (21,22), and mechanisms must be in place to ensure that meiotic cells pass this checkpoint in the presence of unpaired sex chromosomes. This has been extensively studied in mammalian spermatogenesis, where the non-recombining regions of the XY chromosomes fail to pair and remain unsynapsed during meiotic prophase I (23). This leads to the accumulation of repressive chromatin marks and silencing along the two sex chromosomes. The XY chromatin condenses and forms what is called the sex body (XY body), which is inaccessible to transcription machinery (24,25), leading to the complete silencing of both chromosomes. This process is termed meiotic sex chromosome inactivation (MSCI), and it has been described in marsupials, eutherians, nematodes, beetles, chickens, and fruit flies, with the latter two cases being disputed afterwards (26–29,22,30). In the case of fruit flies, a recent study showed that the X chromosome of *Drosophila melanogaster* is not enriched for silencing marks in spermatocytes, suggesting the absence of MSCI (31). Another study used single-cell RNAseq in *Drosophila miranda*, which has the ancestral *Drosophila* X (Muller element A), along with two younger X-linked chromosome arms (Muller AD and Muller C), found that all three X chromosomes have expression patterns consistent with a lack of dosage compensation in late spermatocytes and spermatids (32). The presence of MSCI in some but not all organisms with differentiated XY chromosomes raises the question of what drives it to evolve in the first place, and what alternative mechanisms may be in place in species lacking it.

Currently, the understanding of the sex chromosome specific regulation and the interplay with the tight constraints of gametogenesis is biased towards model species with XY chromosomes. Species with ZW chromosomes (females are ZW, males ZZ) provide an interesting counterpart, as the dosage imbalance and the pairing issues will occur in the female rather than male and during oogenesis rather than spermatogenesis. The status of MSCI in ZW systems is unclear: as after a study reported its presence in the ZW system of chicken during oogenesis (26), another study came to the opposite conclusion (27). A study of MSCI in two Lepidoptera species reported that the Z is euchromatic and transcriptionally active during meiosis (33). However, none of these studies quantified transcriptional output directly, and partial reductions in sex chromosome expression may have been missed. Additionally, as dosage compensation mechanisms evolved to mask the deleterious effects of having a single copy of dosage-sensitive genes, imbalances in the germline of heterogametic females should hypothetically have detrimental effects on oogenesis and early embryogenesis. Due to the limited number of studies on ZW systems, whether dosage compensation is present in the germline cells and whether Z-chromosomes are inactivated during oogenesis are still open questions.

Here, we address these questions using *Artemia* brine shrimp, an aquatic arthropod from the Branchiopoda class with a pair of differentiated ZW sex chromosomes (34). Arthropods have two major types of ovaries: panoistic, where all the germline cells differentiate into oocytes, and meroistic, where only one cell becomes an oocyte and the rest of the germ cells differentiate into nurse cells. Although the majority of crustaceans have panoistic ovaries, meroistic ovaries are typical for Branchiopoda (35). This facilitates drawing parallels between the *Artemia* reproductive system and that of *Drosophila melanogaster*, the most studied insect system in terms of molecular, developmental and morphological data.

In *Artemia* females, oogenesis starts in the two tubular-like ovaries, where the germ cells differentiate into oocytes and up to 70 nurse cells, and where most of the previtellogenesis and vitellogenesis take place (35). *Artemia* nurse cells and oocytes do not show any differences in their morphology until the end of previtellogenesis, where the nurse cells reportedly become polyploid (to increase ribosomal RNA content), and unlike oocytes, do not undergo vitellogenesis and do not produce yolk protein (36). Similar to *Drosophila*, nurse cells remain connected to the oocyte through cytoplasmic bridges and continue supporting it until the end of vitellogenesis, where they are phagocytosed by follicle cells. The oocytes progress through prophase I as they grow in the ovary and move towards the oviduct, where they stay

temporarily. After that, the eggs move to the ovisac and stay there in arrested metaphase I until fertilization (36,37).
Although an analysis in the water flea *Daphnia* (38), the closest model organism to *Artemia*, suggests the sequence conservation of many meiosis genes between insects and crustaceans, their conserved role in crustacean oogenesis and meiosis, and the transcriptional diversity/heterogeneity of cell types in the female reproductive system, have yet to be studied. Here, we create a single nucleus RNA sequencing (RNA-seq) atlas of the *Artemia* ovary, and we identify different somatic and germline cell types, allowing us to perform a detailed comparison with the well-characterized *Drosophila* (39). We further combine RNA-seq and chromatin accessibility (ATAC-seq) data obtained from the same nuclei to investigate the transcriptional and epigenetic dynamics of germ cells during oogenesis. Finally, we characterize the dynamic expression of Z-linked genes in oogenesis, to test whether dosage compensation and sex chromosome inactivation occurs in the germline of the independent ZW system of *Artemia*, enhancing our understanding of sex chromosome regulation during meiosis.

Results

1. snRNA-seq identifies unique cell clusters that share conserved expression programs with *Drosophila*

To resolve the cellular heterogeneity in the *Artemia* female reproductive system and explore the unique regulatory programs and chromatin accessibility in the different cell types, we performed 10x single nucleus RNA sequencing experiments on two biological replicates of pooled ovaries from females kept with males (and therefore putatively mated), and two replicates of 10x Multiome ATAC+Gene expression experiments on pooled ovaries from unmated females (isolated in individual vials after birth to ensure that germ cells would not progress past metaphase I). After preliminary quality checks, and removal of ambient RNA and contamination with cells, we integrated the gene expression data from the four replicates and used dimensionality reduction algorithms to cluster the 20,109 remaining nuclei into 7 clusters (Fig 1A), one of which seems to be specific to the individuals which had access to males (S1 Fig, S1 Table). The clustering resolution (0.05) was chosen based on the specificity-based resolution selection criterion approach (S2 Fig) (40). Heatmaps of the top 10 markers for each cluster identified using Seurat (41) functions (S3 and S4 Figs) suggest that these correspond to functionally differentiated cell types. In order to annotate these distinct cell types, we mapped our clusters to the *Drosophila* ovary dataset from the Fly Cell Atlas (42) using SAMap (43), and filtered for an alignment threshold above 0.2. All the different clusters in *Artemia* map to *Drosophila* clusters (Fig 1B), supporting a high level of conservation of the molecular pathways that define cellular identity in the ovary. Two of our clusters map to germline cells in the *Drosophila* dataset: one to the cells from the germarium region, which we labeled as Germ cells A, and the other to all the later stages of germline cell differentiation, which we refer to as Germ cells B. The expression correlation matrix and its corresponding dendrogram (S5 Fig) show that the clusters mapping to Escort cells and Prefollicle cells are nested with the germ cells, suggesting that those labels might not accurately describe the role of those cell clusters in *Artemia*. To account for this, we defined three major groups (based on the correlation and UMAP distance): group 1 includes Germ cells A, Germ cells B, Escort cells, and Prefollicle cells, group 2 includes Follicle and Tracheal cells, and group 3 includes Ovarian muscle cells.
We also assessed whether the ATAC-seq data contains enough information to disentangle the different cell types in our samples, and whether they would correspond to the cell types inferred from the expression data. We called peaks using MACS2 (44) in each cell type and clustered the nuclei using the peaks. Similar clusters were recovered when using the ATAC

data as from the RNA-seq (Figs 1C and S6), providing further support for their validity.

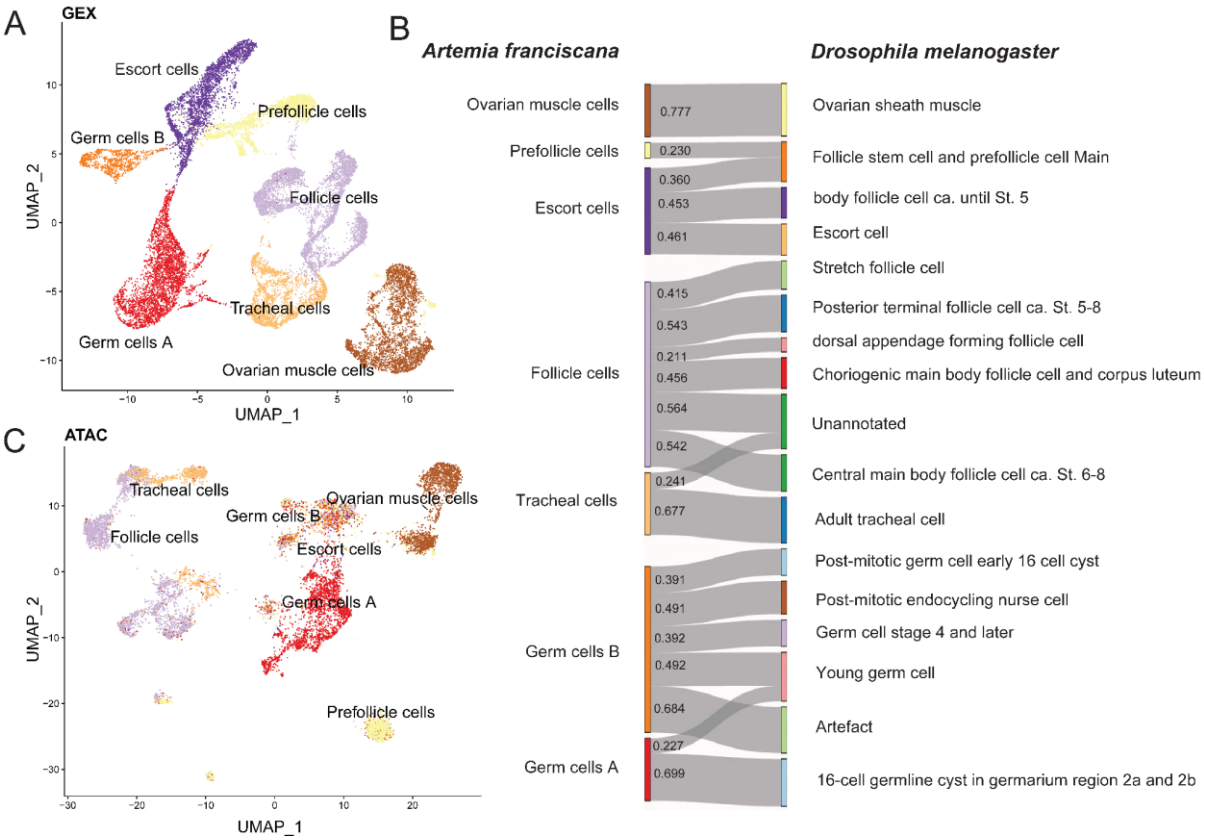


Fig 1: Conserved cell types between Artemia and *Drosophila*. A) Gene expression UMAP. B) Sankey plot showing the mapping between the Artemia and *Drosophila* clusters, the number corresponds to the alignment scores between the clusters. C) UMAP based on the ATACseq data colored based on the gene expression cell-cluster assignments.

2. Germ cells express conserved germline markers and are enriched for meiosis associated genes

To validate the early and later germ cell assignments, we checked the expression of known conserved markers for early (Orb, Fig 2A) and for late (Vas, Fig 2B) germline. Germ cells A show high expression of Orb and Germ cells B show a high expression of Vasa. This is consistent with Germ cells A being an earlier time point in oogenesis than Germ cells B, as Orb is expressed in the region 2 of the germarium in *Drosophila*, very early in oogenesis, and Vasa has been shown to be expressed in the last stages of oogenesis in *Artemia* (45–47). In order to check for the presence of meiotic cells, and to pinpoint the meiotic stages captured in our samples, we also explored the expression of the *Artemia* orthologs of genes known to be involved in the regulation of meiosis in *Drosophila* (48) in each of the clusters (Fig 2C). None of the putative somatic clusters were systematically enriched for any category of meiotic genes. The germ cell A cluster was enriched for *Drosophila* early prophase I genes (pairing and synapsis and double-strand breaks and recombination). The germ cell B cluster was enriched for oocyte maturation (which marks the release from prophase I arrest in *Drosophila*) and germinal vesicle breakdown genes (6), suggesting that this cluster includes late prophase cells. Surprisingly given that spindle formation only occurs in late meiosis, spindle assembly genes were expressed in the early germline, but a similar pattern has been also observed in *Drosophila* (49). It is also important to note that if the nuclear envelope disappears at the end of prophase I (36), the transcriptomes from the stages between the breakdown and reassembly of the nuclear envelope (Telophase I) cannot theoretically be captured with single nucleus sequencing, which likely explains the predominance of prophase cells in our dataset. We also checked the expression of ribosomal proteins, as those were found to be highly expressed in the late stage germline cells of the *Drosophila* ovary single cell Atlas (47). The

Germ cells B cluster has a clear enrichment of those genes, consistent with their assignment as late stage germline cells (Fig 2C).

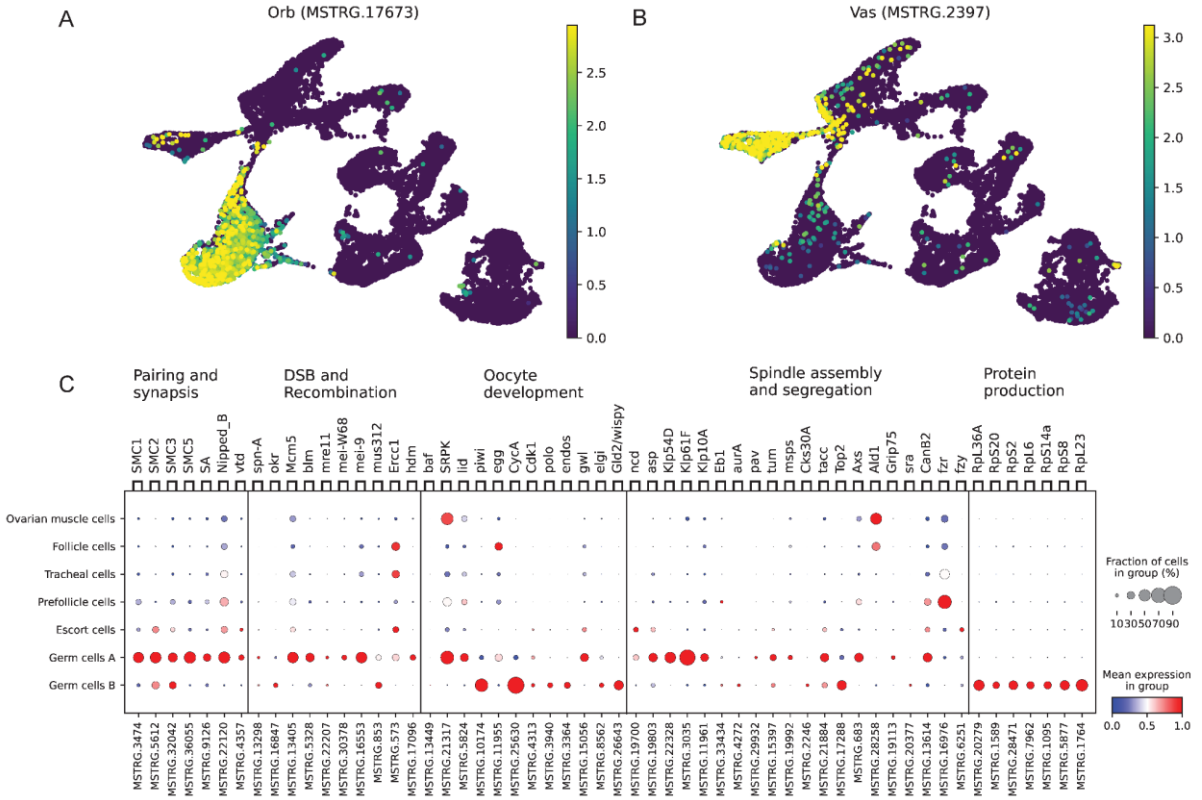


Fig 2: Expression of *Drosophila* germ cell and meiotic markers in *Artemia*. A) Expression of the early germline marker Orb B) Expression of the late germline marker Vasa C) The expression of *Artemia* orthologs of genes involved in the different stages of meiosis in *Drosophila*, along with the expression of genes involved in protein production.

To further explore what pathways may be acting specifically in the germline, we used the hdWGCNA package (50) to perform co-expression network analysis and identify modules (clusters of co-expressed genes) expressed in the different cell types. We constructed networks for the whole dataset and quantified the expression of the modules in the different cell types. The analysis resulted in 14 non-overlapping modules (S7 Fig). We performed differential module eigengene (DME) analysis comparing germline (germ cells A and germ cells B) to all the other clusters and identified 5 modules upregulated in the germline clusters (Fig 3A). Of the 5 upregulated modules in germ cells, 3 had significant PPI enrichment (Figs 3 and S8, S9 and S10). Fig 3 shows modules 7 (269 genes) and 9 (250 genes), which are enriched for chromatin regulation/cell division related terms. Module 7 is highly expressed in Germ cells A and its expression declines across the germline pseudotime (Fig 3B and 3C). Module 9 is expressed in both germ cell clusters (Fig 3D), but peaks in expression in germ cells B. It includes many terms related to histone modifications (methylation and acetylation), further supporting the idea that extensive chromatin remodeling takes place during oogenesis. These modules further support the mitotic/meiotic activity in those clusters and provide novel candidate genes and biological pathways involved in crustacean oogenesis and meiosis.

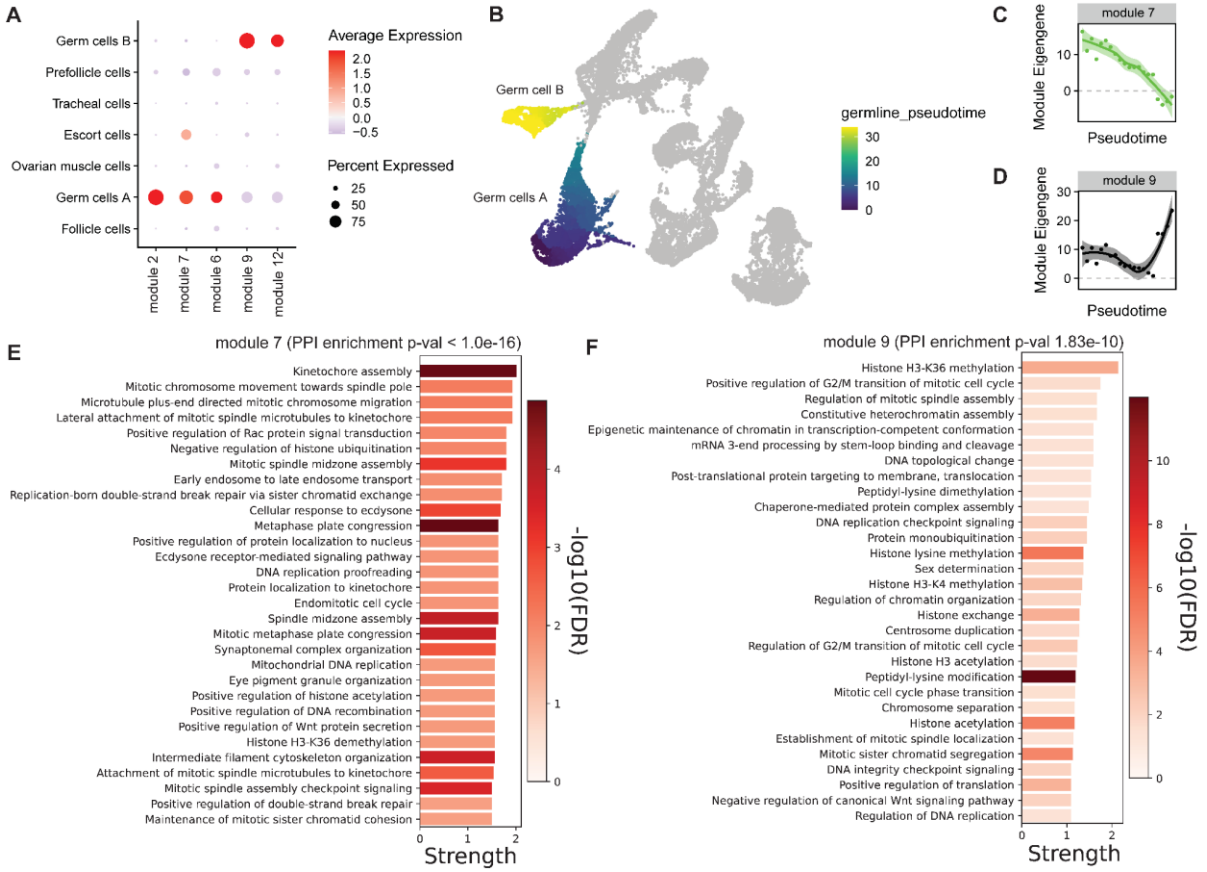


Fig 3: Gene regulatory networks in the germline cells. A) Dotplot depicting the expression of the modules upregulated in germ cells. B) pseudotime trajectory overlaid on the UMAP. C) Module 7 expression dynamics across the germline pseudotime. D) Module 9 expression dynamics across the germline pseudotime. E) Biological process GO enrichment in module 7. F) Biological process GO enrichment in module 9. Strength (retrieved from StringDB) is estimated as the log10(number of observed proteins with a term/ number of expected proteins with the term in a random network of the same size).

3. Transcription is repressed in late germ cells

During meiotic prophase, the chromatin of *Drosophila* oocytes condenses, and this condensation is accompanied by transcriptional repression (48,51). We explored potential signatures of such changes in our gene expression and ATAC data (replicates 3 and 4). The number of ATAC-seq counts in peaks in the Germ cells B cluster is very low compared to all the other clusters (Fig 4B, number of fragments per cluster is shown in S11 Fig), consistent with the chromosomes being highly condensed during late prophase/metaphase I. The nuclei assigned as Escort cells also show low RNA counts, likely reflecting the fact that they are misassigned Germ cells A and B (as this cluster is largely missing from our multiomics dataset). To further characterize the transcriptional activity of the different clusters, we estimated the percentage of spliced and unspliced transcripts in the different clusters using Velocyto with the raw data. We observed a high percentage of spliced RNA in the Germ cells B cluster compared to all the other clusters (Fig 4C, p=1.35e-08, Chi-square contingency test). Only replicates 3 and 4, for which ATAC-seq data were available, are depicted in Fig 4C, but replicates 1 and 2 also show the same pattern (S12 Fig). The higher rate of spliced RNA can be linked to a decrease in or a complete pause of transcriptional activity, in line with the observed reduction in chromatin accessibility. While such a pattern has not been reported in similar datasets, running Velocyto on the Fly Cell Atlas ovary data also yielded a much higher percentage of spliced transcripts in the young germ cells cluster of *Drosophila* than in other cell types (S13 Fig). In order to explore the dynamics and time of onset of this putative transcriptional silencing, we performed pseudotime analysis using germ cells of replicates 3 and 4 (nuclei assigned as Escort cells in those replicates were included as they appear to be

wrongly clustered Germ cells A and B)(Fig 4D). We see that both the ATAC and the RNA counts decrease across the germline pseudotime, suggesting that transcriptional repression is progressively established in the Germ cells A cluster (Fig 4E and 4F). The RNA counts seem to rise towards the end the germline pseudotime, which could be a sign of transient transcriptional reactivation, similar to what happens in *Drosophila* oocytes during oogenesis, between stages 9 and 11 (prophase I arrest ends at stage 13)(9). However, as the ATAC counts do not show a similar pattern, the pattern may not be biological.

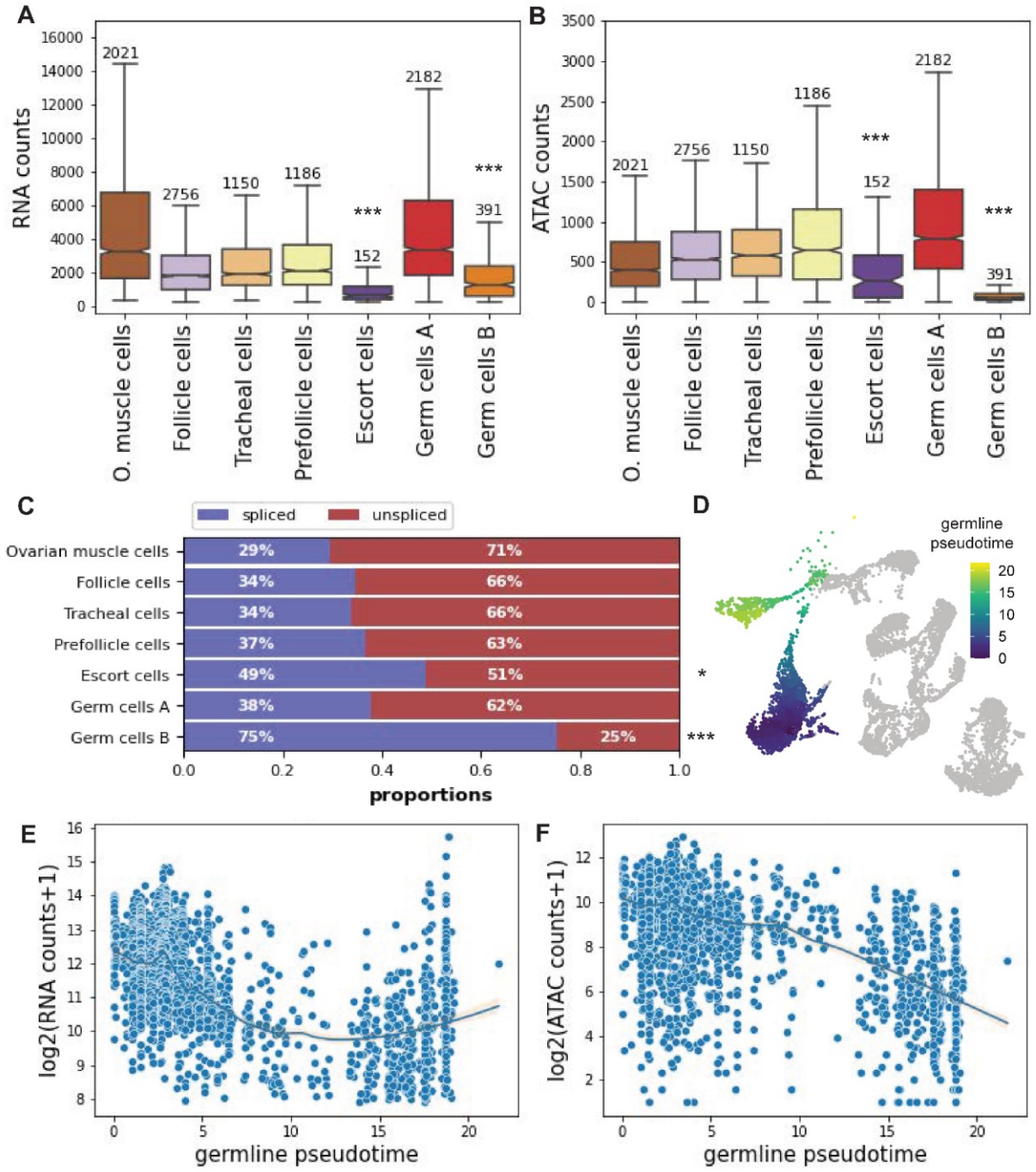


Fig 4: Unique features of germ cells B. A) The snRNA-seq counts per cell in the different clusters (replicates 3 and 4, number of nuclei above each boxplot). The stars show the significance values for group 3 and somatic clusters comparisons (wilcoxon rank-sum test). B) The snATAC-seq counts (in peaks) in the different clusters (replicates 3 and 4, number of nuclei above each boxplot). The stars show the significance values for group 3 and somatic clusters comparisons (wilcoxon rank-sum test). C) The percentage of spliced and unspliced transcripts (replicates 3 and 4). The stars show the significance values for group 3 and somatic clusters comparisons (Chi-square contingency test). For the stars, *** denotes p-value <= 0.001, ** denotes p-value <= 0.01, and * denotes p-value <=0.05. D) pseudotime analysis (Germ cells A, Germ cells B and escort cells) using replicates 3 and 4 E) RNA counts in Germ cells A, B, and escort cells across the germline pseudotime (the line is deptics the local regression result with confidence intervals) F) ATAC counts in germ cells A, B, and escort cells across the germline pseudotime (the line depicts the local regression result with confidence intervals)

4. Downregulation of the Z chromosome in the germline

After identifying the different cell types in our dataset, we explored the expression of the differentiated region of the Z chromosome (A previously identified ~13 MB Z-linked region with half the coverage in females compared to males, which we refer to as S0 (52)) to assess whether meiotic sex chromosome inactivation is present in *Artemia*. Fig 5A shows the inferred S0/Autosomal values per nucleus on the UMAP, and highlights a strong excess of cells where the S0 seems to be downregulated in the two germ cell clusters. To explore this pattern quantitatively, we estimated the ratio of the mean expression of the genes in the S0 region (446 genes) and mean expression of Autosomal genes (26439 genes). The boxplot of the S0/Autosome expression shows that the Z-specific region is indeed downregulated in the germline cell clusters compared to somatic cells ($p<10\text{-}16$ with wilcoxon rank-sum test; Clusters belonging to groups 2 and 3 are used as the somatic control). We recover the same S0/Auto patterns when using the mean of non-overlapping bins of 446 autosomal genes instead of the mean of all autosomal genes (See methods section: “Estimation of S0/Autosomal ratio using non-overlapping autosomal windows” and S14 Fig).

The median of the expression ratio in Germ cells B is around 0.5, which is overall more consistent with lack of dosage compensation than with true Z inactivation. In order to check if any cells have expression patterns consistent with additional repression of the S0, we classified all cells in each cluster into bins of decreasing S0:A expression ratio that should reflect the presence of dosage compensation and/or repression of the S0 (S0/Auto ratio for all cells is adjusted by adding 1-median(S0/Auto of somatic clusters)):

- Complete or partial dosage compensation: S0/Auto > 0.66
- Lack of dosage compensation: S0/Auto <= 0.66 and S0/Auto > 0.33.
- Repressed: S0/Auto <= 0.33.

As expected, the vast majority of cells of the somatic clusters have full dosage compensation, with virtually none being classified as Z-repressed. Germ cells A are enriched for lack of dosage compensation (27.83%, p=0.0003 with Chi-square contingency test). Germ cells B show a high enrichment for lack of dosage compensation (55.24%, p=1.07e-12), but we also observe a high proportion of cells with expression consistent with repression of the differentiated region (16.79%, p=0.0002 compared with autosomal clusters). As an additional measure, we used percentile-based cutoffs to control for the heterogeneity of Z chromosome regulation status caused by noise, and we recovered the same enrichment patterns (See methods section “Z-chromosome regulation status using percentile-based cutoffs” and S15 Fig).

While these results point to a small subset of germline cells (less than 20% of the cells in the cluster) showing at least partial Z-inactivation, it is notoriously difficult to fully exclude that absence of dosage compensation, along with sparse and noisy data, could create such a pattern. To bypass this, we reasoned that absence of dosage compensation should only affect the S0 region of the Z, which no longer has W-homologs, whereas both younger non-recombining but undifferentiated regions (S1 and S2), as well as the pseudoautosomal region, should not be affected. On the other hand, depending on the mechanism at play, inactivation may spread to other regions of the chromosome. We therefore explored the expression of the other regions of the Z-chromosome, which include the pseudoautosomal region (PAR), and the younger strata S1 and S2 (52)(S16 Fig). Both germline clusters show lower expression levels of these undifferentiated regions compared to somatic clusters, with a more consistent downregulation in Germ cells B.

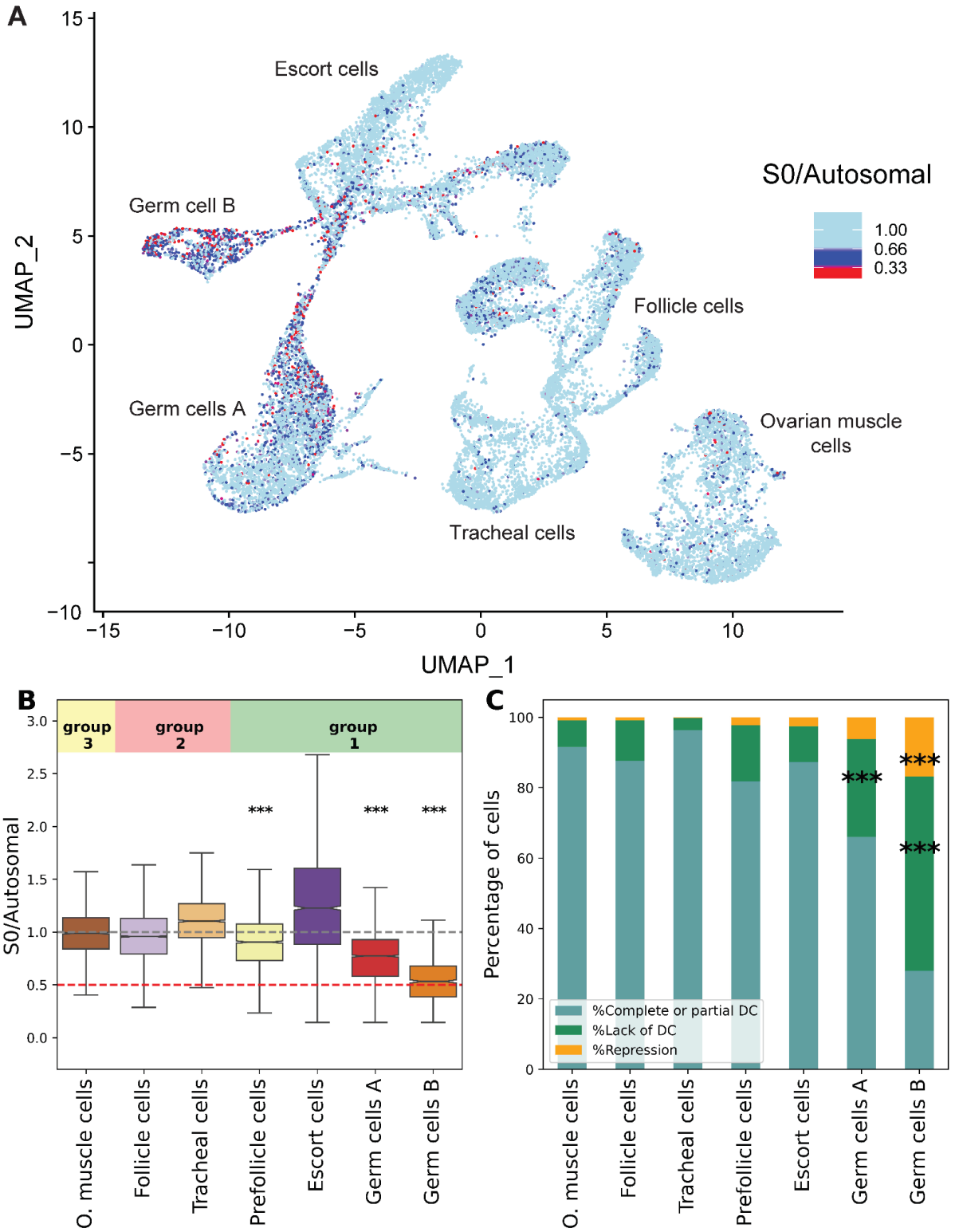


Fig 5: Downregulation of the Z chromosome in the germ cells. A) UMAP showing the log2(S0/Autosomes) expression per cell. B) The mean(S0)/mean(Autosomes) expression per cell estimated using the normalized counts matrix. The stars show the significance values for group 3 and somatic clusters comparisons (wilcoxon rank-sum test). C) The percentage of cells that are dosage compensated (DC), lack dosage compensation (Lack of DC) and repressed (Repression). The stars show the significance values comparing group 1 clusters and somatic clusters (%Lack dosage compensation vs rest, and %Repression vs rest using Chi-square contingency test). For the stars, *** denotes p-value <= 0.001, ** denotes p-value <= 0.01, and * denotes p-value <=0.05.

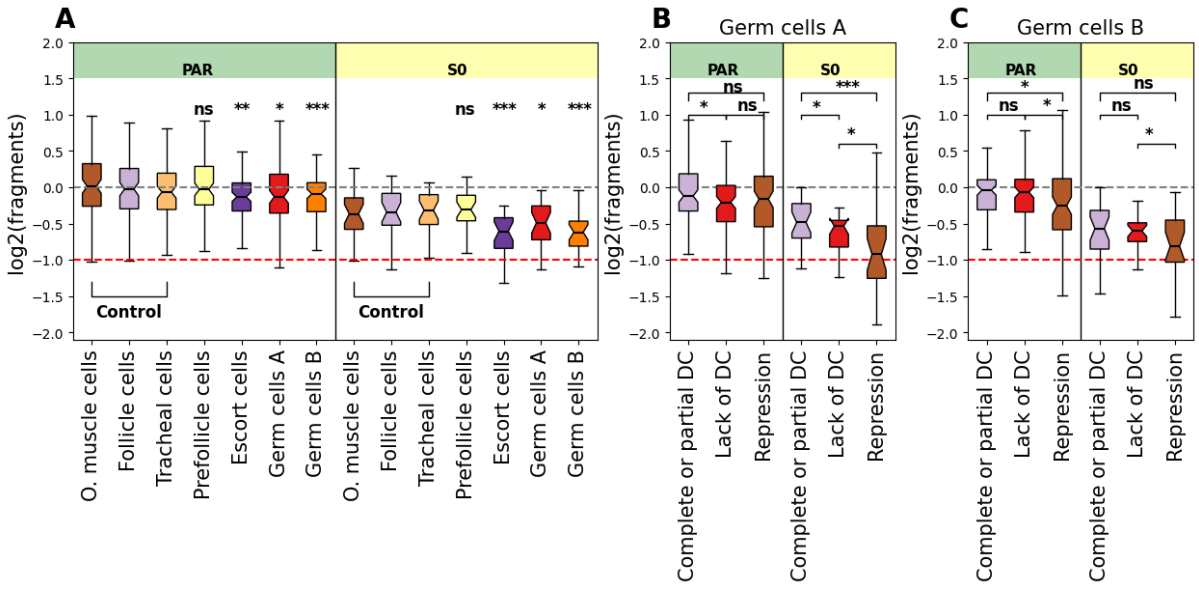


Fig 6: Decreased Z-chromosome accessibility in the germline clusters. A) log2(fragments) for autosomal, PAR, and S0 in windows of 500,000 bp estimated from pseudo-bulks of all nuclei within a cluster adjusted by subtracting the median(log2(autosomal windows)) of each cluster. The stars show the significance values for the comparisons between group 3 clusters and somatic clusters (wilcoxon rank-sum test). B) and C) log2(fragments) for autosomal, PAR, and S0 in windows of 500,000 in Germ cells A and B split into pseudo-bulks based on the expression zone of the Z chromosome adjusted by subtracting the median(log2(autosomal windows)) of each cluster. The stars show the significance values for comparisons between the different categories (wilcoxon rank-sum test). For the stars, *** denotes p-value <= 0.001, ** denotes p-value <= 0.01, and * denotes p-value <=0.05. The red line is at -1, and corresponds to a two fold decrease in the number of fragments.

Finally, to explore whether the downregulation of the Z chromosome in germ cells corresponds to a change in chromatin accessibility, we pooled the counts from each cluster together and counted the number of ATAC fragments in the differentiated and pseudoautosomal regions of the Z in windows of 500,000 bp and compared them to the number of fragments in autosomal windows (Fig 6A). We see a slight decrease in accessibility of the two germline clusters compared to the somatic clusters in the S0 (p-value of 0.018 and 4.55e-06 for Germ cells A and B clusters respectively) and PAR (p-value of 0.037 and 0.00018 for Germ cells A and B clusters respectively) regions, which suggests that the Z chromosome is less accessible than the autosomes. As dosage compensated cells may be masking the signal when pooling all the counts together, we split the cells based on their Z expression zone (Fig 6B and 6C). Although most of the comparisons are not significant, we still see that the number of fragments is consistent with the expression zone, with the repressed cells having the lowest number of fragments originating from the S0 and PAR regions.

Discussion

- **Conserved cell identity programs across Arthropoda**

In this study, we used single nucleus RNA sequencing to resolve the cellular complexity of the *Artemia* ovary. We identify clusters of cells with distinct expression patterns and show that they share different levels of expression-based homology with the *Drosophila* ovarian clusters from the Fly cell atlas. This suggests that many of the expression programs that give rise to cellular identity and function are highly conserved despite ~505 MYA (CI: 474.8 - 530.0 MYA) of divergence between Crustacea and Insecta (53). Ovarian muscle cells show the highest level of conservation, in line with a previous study that showed a high similarity in orthologous gene expression of muscle cells across several vertebrate and invertebrate species (54). Two clusters, which show some expression similarity to each other (S5 Fig), map to Tracheal and Follicle cells in *Drosophila* (epithelial cell types). In *Drosophila*, Tracheal cells form the tracheal system, which transports oxygen to the different organs (55), while follicle cells are involved in many aspects of oogenesis, including control of egg shape, eggshell formation, and

formation of appendages, such as the dorsal appendage (56). The two identified clusters (in addition to prefollicle cells) express trachealess (trh, S17A Fig), which is essential for the initiation and maintenance of invagination (57). The trachealess ortholog has also been found to be expressed in the salt gland of nauplii and in the thoracic epipod of juveniles in *Artemia franciscana, s*uggesting it might play other roles in this species (58).

Early germ cells (germ cells A) also show a high level of conservation between the two species. Previous studies have linked the observed conservation of several aspects of germline cyst development between distant species, such as of mice and *Drosophila*, to the critical role of germ cells in the preservation of the nuclear genome and the importance of early oocytes for embryonic development (59)**.** We identify a single cluster in our dataset that maps to all the late stage germline clusters in *Drosophila*, including nurse cells. This is in line with expression patterns in this species, where all late germline and nurse cell clusters are highly correlated (S18 Fig). It should therefore be noted that our germline clusters may also contain a mix of developing oocytes and closely related nurse cells.

Some of the cluster annotations should be interpreted more cautiously, as they likely represent cell types with less conserved transcriptional programs, making inferences based on *Drosophila* less reliable. The *Artemia* cluster mapping to Escort cells is nested with the two germline clusters in the dendrogram, unlike Escort cells in the *Drosophila* dendrogram, which are nested with somatic cells (S5 Fig). The absence of Escort cells from the unmated females suggests those cells are either late stage follicle cells, late stage germline/embryonic cells, or sperm (the minority of cells that are assigned to the same cluster in replicates 3 and 4 after integration are likely to be misclustered germ cells A or B). We did not find any enrichment of testis biased genes in any of the different clusters, which rules out contamination by sperm (S19A and S19 Fig), but could not distinguish between other possibilities. The cluster mapping to the *Drosophila* prefollicle cells had the lowest alignment rate (0.230), and it also shows some expression similarity to the germ cell clusters (S5 Fig), suggesting the assignment as follicle cells may not be fully accurate. However these cells have high expression of phagocytosis and apoptotic cell clearance genes *draper* and *draper-like* (S17D Fig), which are expressed in follicle cells to promote nurse cell death in *Drosophila* (60).

- **The unusual regulation of the Autosomes and Z chromosome in germline cells**

We use 10x single cell RNAseq and ATACseq to explore the expression and chromatin accessibility changes during female oogenesis in *Artemia* brine shrimp. We observe a dramatic decrease in the ATAC counts in the Germ cells B cluster, along with a noticeable decrease in the total RNA counts and the percentage of unspliced RNA. Those observations further support the idea that this cluster consists at least in part of late prophase cells, and the lack of ATAC signal is likely due to the compaction of chromatin and the establishment of prophase I arrest. A similar pattern has been observed in *Drosophila*, where the total number of ATAC peaks decreases dramatically in the later stages of oogenesis (whole ovaries) compared to GSCs and young ovaries (51). In mouse single cell data, the number of ATAC peaks in mitotic cells decreases dramatically in the progression from prophase I to metaphase I (61). The compact structure of the chromatin during mitotic and meiotic prophase is thought to present a barrier to many transcription factors, which causes a reduction in the levels of gene expression (9,62). In meiosis, the global silencing of transcription in oocytes is highly conserved, and the extensive remodeling of the oocyte chromatin seems to play an important role in the oocyte to embryo transition (63). We also checked the expression of three genes that have been shown to play an important role in the chromatin remodeling of the *Drosophila* oocyte (their knockdowns introduced significant disturbances to the oocyte epigenome)(9). Lid, which is associated with the activating histone mark H3K4me3 is expressed in germ cells A, and Ash1 and Bap1, associated with the repressive histone mark H3K27me3, are expressed in Germ cells B (S17E Fig).

In many species, gametogenesis coincides with the loss of dosage compensation. In the case of female mammals, this takes the form of reactivating the silenced X in the germline cells (18,19). In *Drosophila* males, the lack of dosage compensation manifests in the absence of X chromosome upregulation in primordial germ cells, spermatocytes and spermatids (16,17,64). This could be the result of the global reprogramming of the epigenome required for the

generation of a “clean slate” for transmission to the embryo. How such a clean slate is achieved in the presence of ZW chromosomes is unclear, as loss of dosage compensation in oocytes could lead to imbalances in expression that are then transmitted to the embryo maternally (65). In our data, the two germline clusters (germ cells A and germ cells B) show lower S0/Autosomal expression (Fig 5B), which seems to be driven by the enrichment in cells with S0 downregulation, consistent with a lack dosage compensation compared to somatic clusters. The ATACseq results show a similar pattern, where the germline cells have fewer counts in the S0 region compared to the somatic clusters (Fig 6). Other mechanisms must therefore be in place to avoid imbalances in the expression of maternal RNAs, such as their production by compensated nurse or follicle cells.

- **Z chromosome repression in the germline**

We find that the two germline clusters include cells which seem to have very low S0 expression (consistent with repression), and when we split the ATAC counts based on the expression zones, the repressed cells seem to have lower counts in the Z-specific region, consistent with the lowest accessibility. The fact that the whole Z-specific region seems to be downregulated and less accessible suggests that a whole chromosome mechanism may be in action, reminiscent of meiotic sex chromosome inactivation. Additionally, lack of dosage compensation in other species seems to result in less than 2 fold decrease in the expression of X/Z-linked genes (~1.5 in the *Drosophila* testes and ~1.6 in the chicken gonads) (64,66–68). In our analysis, the distribution of S0/Autosome ratios per cell in the germ cells B cluster is centered at 0.5 (2-fold decrease in expression). If one assumes an expected value of 0.66 for lack of DC, then the observation of 0.5 might suggest a combination of lack of DC and sex chromosome repression. It is important to note some limitations of our data, including the low capture level of total mRNA per cell (high dropout rate), high ambient RNA, and sparse read mapping, which make confident inferences of silencing difficult. We used the same approach and percentile-based thresholds to check whether we see a similar pattern in the *Drosophila* testes dataset (69), and we only observe an enrichment in cells lacking dosage compensation in some of the germline clusters (mainly in the meiotic and post-meiotic cell types), but no cells show extreme repression of the X chromosome (S20 Fig). Additionally, we explored the expression of the *Artemia* genes annotated under the ‘Facultative heterochromatin assembly’ GO and ‘Constitutive heterochromatin assembly’ and the majority show Germ cells B specific expression pattern (S17B and S17C Fig). Taken together, our data therefore generally points towards the possibility that repression of the sex chromosome occurs during oogenesis, although a demonstration that repressive chromatin marks are present on the Z will be needed to confirm this.

The evolutionary hypotheses stemming for the mammalian case of MSCI sparked a lot of interest in understanding the conditions that favored the evolution of such a mechanism, and whether it is a universal consequence of having heteromorphic sex chromosomes. The fact that the reports of MSCI in *Drosophila* and chicken have been disputed later, and in many other species, such as moths and butterflies, the evidence so far suggests its absence, implies that the mechanism is either not as universal as initially assumed or that those species are exceptions to the rule. In particular, both *Drosophila* males and *Lepidoptera* females have achiasmatic meiosis (70,71), and the chicken ZW chromosomes achieve complete heterologous synapsis (29). *Artemia* have similar recombination rates in males and females, arguing against achiasmy in females, perhaps providing an explanation for why meiotic sex chromosome silencing may have been favored. More generally, broader sampling is needed to understand the role of the sex chromosome system (female or male heterogamety), the extent of sex chromosome differentiation (homomorphic or heteromorphic), the meiotic idiosyncrasies (type of pairing and presence or absence of recombination), and repeat content/meiotic driver presence/activity in promoting the evolution of MSCI. Our study of meiotic sex chromosome regulation in a female heterogametic system with a well differentiated region is a step in this direction. Our work also highlights single-nucleus RNA sequencing as a useful alternative to traditional approaches, such as epigenetic profiling and RNA-FISH, for identifying promising models for the study of meiotic sex chromosome regulation in species where it is difficult to isolate/identify nuclei of meiotic cells.

Methods

Single-nucleus sequencing of the *Artemia* Female reproductive system

We isolated *Artemia franciscana* adult females from either a colony or from vials where they were kept individually (see below), and washed them in milliQ water to remove any excess salt. The ovaries were dissected in ice-cold DPBS and then moved to an eppendorf with DPBS and placed on ice. The sample was then washed once with DPBS, and after spinning down, the DPBS was removed without disrupting the pellet of ovaries, and 1 ml of the homogenization buffer was added to the sample. Following the protocol described in (72), the whole contents of the eppendorf were then transferred to a 1 ml dounce homogenizer. The nuclei were then released by 20 strikes with the loose Dounce pestle and 40 strikes with the tight pestle on ice. The sample was then filtered through a 35 um cell strainer into a FACS tube, and then filtered again using a 40 um Flowmi cell strainer into a 1.5 ml eppendorf. Each sample was then centrifuged for 10 minutes at 4°C and 1000 g. The supernatant was discarded and the pellet was resuspended using ~300–500 µL of resuspension buffer. For the 10x Multiome samples, 10 µL/mL of 0.5% Digitonin (BN2006, Invitrogen) was added to the homogenization buffer to permeabilize the nuclei and facilitate the access of the tagmentation enzyme to the chromatin; the samples were incubated in the buffer for 5 minutes after homogenization before proceeding with the remaining steps. The samples were transferred for 10x genomics sorting and sequencing at the Vienna BioCenter Next Generation Sequencing (NGS) Core Facility. In all the replicates, 16,000 nuclei were loaded on the chip, targeting 10,000 individual nuclei.

For the 3' GEX experiments, 25 mated females were used in each replicate, and for two replicates the 10x Multiome ATAC+Gene expression experiments, the same number of unmated females (isolated at the Naupliar stage and maintained in individual vials until they reached sexual maturity) were used per replicate. As our experiments include mixed genotypes, it was possible to estimate the percentage of ambient RNA in each replicate using Souporcell (73)(Supplementary table 2).

Preprocessing, Quality Control, and Integration of the different replicates

The reads from each sample were mapped to the *A. franciscana* genome (52), annotated using StringTie2 (74), using 10x Genomics Cell Ranger 5.0.0 for the two 3'GEX samples and using Cell Ranger ARC 2.0.2 for the two Single Cell Multiome ATAC + Gene Expression samples (75,76). The CellBender v0.2 (77) package was run on the raw gene-by-cell matrix from each replicate to remove the technical artifacts and background noise and produce an improved estimate of gene expression per cell. Specific low count thresholds, droplet training fractions, and false positive rates were chosen for each sample following the Cellbender best practices (https://cellbender.readthedocs.io/en/latest/troubleshooting/), and are provided in the git page listed below. The output of CellBender was then loaded into Seurat (41), where nuclei with < 10 features, nuclei with > 3% mitochondrial content, and doublets were removed. The filtered nuclei from all the replicates were then loaded into Seurat, and only cells with (nFeature_RNA > 200 & nFeature_RNA < 25000) were retained. The highly variable features were identified using DUBStepR (78) with default parameters and the replicates were integrated using harmony, clustered using graph-based approaches, and then visualized using non-linear dimensionality reduction UMAP. The resolution for clustering (0.05) was determined using the marker specificity-based analysis from scMiko (40). The cluster markers were identified using two different Seurat functions: FindConservedMarkers and FindAllMarkers. As FindConservedMarkers identifies the differentially expressed genes between the clusters which are conserved across the replicates, we reasoned that the results would not be reliable in the case of Escort cells due to their absence from replicates 3 and 4. Therefore, we also used FindAllMarkers, which does not take the replicate information into account. To ensure that our results are not an artifact of ambient RNA removal, we performed the analysis with the raw counts. The global structure is preserved (S21 Fig), along with the significant differences between the identified dosage compensated, not dosage compensated, and repressed nuclei (S22 Fig). Additionally, despite the noisiness of the raw data, the

enrichment in Orb and Vas was clear in the germline cells compared to the somatic clusters (S21B and S21C Fig).

ATAC-seq clustering and Analysis

For the clustering analysis, the raw count matrix was loaded into Seurat and filtered to keep only the cells that are in the expression clusters. The peaks were called per cluster using MACS2 (44) in each replicate separately and the resulting peaks were then combined. The data was then normalized using RunTFIDF (method=3), and the variable features were identified using FindTopFeatures with min.cutoff = 'q3'. RunSVD was then used to perform latent semantic indexing (LSI) and the nonlinear dimensionality reduction was performed using UMAP. The same clustering as for the gene expression analysis was used for visualization. We have also checked the correlation between the peaks and expression. We divided the genes into three categories (<=20% for low expression genes, >20% and <80% for medium expression, and high expression genes >=80%). We used the mean of the ATAC counts in the linked peaks for each gene within a cluster, and S23 Fig shows that the peak enrichment corresponds to the expression level in all the clusters (Germ cells B and Escort cells do not show as clean a pattern due to the low number of peaks detected).

Integration of the *Artemia* atlas with the *Drosophila* fly cell atlas

We used the SAMap blast-based mapping script to map the *Artemia* transcripts to the *Drosophila* cds (dmel-all-CDS-r6.31.fasta) downloaded from flybase (79) and filtered to keep only the longest isoform for each gene. The *Drosophila* single nucleus data (10x VSN Ovary (Stringent), 10x genomics, H5AD) was downloaded from: https://cloud.flycellatlas.org/index.php/s/zgZe3Zsegpn5Bpg/download/s_fca_biohub_ovary_10x.h5ad. SAMap (43) was then run using the jupyter notebook provided in the supplementary material. An alignment threshold of 0.2 was used for displaying the cluster correspondence using the sankey diagram.

Identification of meiosis and germline markers

The germline and meiosis markers were found in the literature (48,47) and the *Artemia* homologs were identified as the reciprocal best hits based on the SAMap mapping output. The expression of the markers in *Drosophila* is shown in S24 Fig.

Networks analysis using hdWCGNA

In order to construct the co-expression network, we ran hdWCGNA (50) on genes expressed in at least 5% of the nuclei. We constructed the metacells grouping by the cell type and replicate information, and we constructed the co-expression network for all the clusters simultaneously. We performed differential module eigengene (DME) analysis comparing the germ cells A and germ cells B group to a group made of all the other clusters. We then performed pseudotime trajectory analysis on the whole dataset, isolated the germline cells (Figs 3B and S25) and explored the module dynamics across the pseudotime (Figs 3C and 3D and S26).

Quantifying the proportion of spliced and unspliced transcripts

We used Velocyto (80) to annotate spliced and unspliced transcripts and generate spliced/unspliced count matrices for each replicate using the output from Cell Ranger. We then used SCANPY (81) to merge the matrices, and scVelo (82) to plot the proportions of spliced/unspliced counts (jupyter notebook provided on the gitpage). For the *Drosophila* estimation, we downloaded the raw ovary Fastq files, aligned them to the *Drosophila* genome (Drosophila_melanogaster.BDGP6.32.dna.toplevel.fa) using 10x Genomics Cell Ranger 5.0.0, and used Velocyto to get the spliced and unspliced counts, and then merged the matrices with the expression matrix provided on the Fly cell Atlas (scripts provided on gitpage).

Protein Interaction network and GO enrichment analysis

We translated the *Artemia franciscana* transcriptome generated using StringTie2 with the perl script GetLongestAA_v1_July2020.pl, and the translated sequences were uploaded to https://string-db.org/, where the PPI and GO enrichment analyses for the modules were performed. The annotated proteome is accessible using the following link: https://version-12-0.string-db.org/organism/STRG0A95DBT.

Z-chromosome regulation status using percentile-based cutoffs

As the within cluster variation in the status of Z-chromosome expression is possibly driven by noise, we implemented more conservative thresholds that apply a 5% false positive rate to the first category in each comparison to provide a noise-sensitive estimate of the cluster-specific enrichments (S15 Fig):

- Complete or Partial dosage compensation: S0/Auto > 5th percentile of S0/Auto in somatic clusters (>0.57).
- Lack of dosage compensation: S0/Auto <= 5th percentile of S0/Auto in somatic clusters and S0/Auto > (5th percentile of S0/Auto in somatic clusters)/2 (<=0.57 and >0.28).
- Repressed: log2(S0/Auto) <= (5th percentile of S0/Auto)in somatic clusters)/2 (<=0.28).

Estimation of S0/Autosomal ratio using non-overlapping autosomal windows

To ensure that our S0/Autosomal expression estimates are not affected by the low expression throughout the genome, as is the case for some germline cells (see results section 3), we divided the genome into 49 non-overlapping sliding windows with the same number of genes as the S0 (446 genes). We reasoned that regions of similar gene counts as the S0 are as susceptible to the low detection rates and non-biological zeros that affect single-cell RNAseq data, and can therefore be used to ensure the overall patterns are not technical artifacts. In S14 Fig, we show the distribution of the per cluster medians of S0/Autosomal window for all the 49 windows.

Data and Code Availability

The scripts used in the analysis can be accessed on the GitHub page: https://github.com/Melkrewi/Artemia-snRNAseq-Project. The raw data are available on the NCBI short read archive (BioProject number PRJNA1128544). The Seurat objects produced in the analysis, the loom files generated by Velocyto, the reciprocal best hits used for the SAMap analysis, and the module tables and their GO enrichment results are available for download on ISTA Research Explorer (ISTA REx) using the following link: https://doi.org/10.15479/AT:ISTA:17362. The single nucleus atlas can be viewed with this link on the UCSC Cell Browser: https://brine-shrimp-repro.cells.ucsc.edu.

Funding

Funding: This research was funded by the Austrian science fund (FWF), as part of the SFB Meiosis consortium (https://sfbmeiosis.org/, grant ID FWF SFB F88-10) to BV. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank the Vicoso group for their valuable comments on the earlier draft of the manuscript. We would also like to thank the Vienna BioCenter Next Generation Sequencing (NGS) facility staff, and in particular, Thomas Grentzinger for his support with the handling and sequencing of the samples, the scientific computing unit at ISTA for the computational resources, Brittney Wick for the help with hosting our data on the UCSC Cell Browser, and Lora B. Sweeney for her valuable input at the different stages of the project.

References

1. Gilbert SF. Oogenesis. In: Developmental Biology 6th edition [Internet]. Sinauer Associates; 2000 [cited 2023 Dec 5]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK10008/
2. Lenormand T, Engelstädter J, Johnston SE, Wijnker E, Haag CR. Evolutionary mysteries in meiosis. Philos Trans R Soc B Biol Sci [Internet]. 2016 Oct 19 [cited 2024 Feb 14]; Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0001
3. Loidl J. Conservation and Variability of Meiosis Across the Eukaryotes. Annu Rev Genet. 2016;50(1):293–316.
4. Zickler D, Kleckner N. Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harb Perspect Biol. 2015 Jun;7(6):a016626.
5. Hajkova P. Epigenetic reprogramming in the germline: towards the ground state of the epigenome. Philos Trans R Soc B Biol Sci. 2011 Aug 12;366(1575):2266–73.
6. Stetina JRV, Orr-Weaver TL. Developmental Control of Oocyte Maturation and Egg Activation in Metazoan Models. Cold Spring Harb Perspect Biol [Internet]. 2011 Oct [cited 2023 Dec 18];3(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3179337/
7. Hinnant TD, Merkle JA, Ables ET. Coordinating Proliferation, Polarity, and Cell Fate in the *Drosophila* Female Germline. Front Cell Dev Biol [Internet]. 2020 [cited 2024 Feb 15];8. Available from: https://www.frontiersin.org/articles/10.3389/fcell.2020.00019
8. Lei L, Spradling AC. Mouse oocytes differentiate through organelle enrichment from sister cyst germ cells. Science. 2016 Apr 1;352(6281):95–9.
9. Navarro-Costa P, McCarthy A, Prudêncio P, Greer C, Guilgur LG, Becker JD, et al. Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. Nat Commun. 2016 Aug 10;7(1):12331.
10. Samata M, Alexiadis A, Richard G, Georgiev P, Nuebler J, Kulkarni T, et al. Intergenerationally Maintained Histone H4 Lysine 16 Acetylation Is Instructive for Future Gene Activation. Cell. 2020 Jul 9;182(1):127-144.e23.
11. F Z, E L, R S, F K, O B, N I. Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Science [Internet]. 2017 Jul 14 [cited 2024 Mar 9];357(6347). Available from: https://pubmed.ncbi.nlm.nih.gov/28706074/
12. Inoue A, Jiang L, Falong L, Suzuki T, Zhang Y. Maternal H3K27me3 controls DNA methylation-independent genomic imprinting. Nature. 2017 Jul 27;547(7664):419–24.
13. Conti M, Hsieh M, Zamah AM, Oh JS. Novel signaling mechanisms in the ovary during oocyte maturation and ovulation. Mol Cell Endocrinol. 2012 Jun 5;356(0):65–73.
14. Deady LD, Sun J. A Follicle Rupture Assay Reveals an Essential Role for Follicular Adrenergic Signaling in *Drosophila* Ovulation. PLOS Genet. 2015 Oct 16;11(10):e1005604.
15. Disteche CM. Dosage compensation of the sex chromosomes and autosomes. Semin Cell Dev Biol. 2016 Aug;56:9–18.
16. Witt E, Shao Z, Hu C, Krause HM, Zhao L. Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in *Drosophila* testis. PLOS Genet. 2021 Aug 17;17(8):e1009728.
17. Ota R, Hayashi M, Morita S, Miura H, Kobayashi S. Absence of X-chromosome dosage compensation in the primordial germ cells of *Drosophila* embryos. Sci Rep. 2021 Mar 1;11:4890.
18. Sangrithi MN, Turner JMA. Mammalian X Chromosome Dosage Compensation: Perspectives From the Germ Line. BioEssays. 2018;40(6):1800024.
19. Mattimoe T, Payer B. The compleX balancing act of controlling X-chromosome dosage and how it impacts mammalian germline development. Biochem J. 2023 Apr 26;480(8):521–37.
20. Heijden GW van der, Eijpe M, Baarends WM. The X and Y chromosome in meiosis: how and why they keep silent. Asian J Androl. 2011 Nov;13(6):779.
21. Turner JMA, Mahadevaiah SK, Fernandez-Capetillo O, Nussenzweig A, Xu X, Deng CX, et al. Silencing of unsynapsed meiotic chromosomes in the mouse. Nat Genet. 2005 Jan;37(1):41–7.

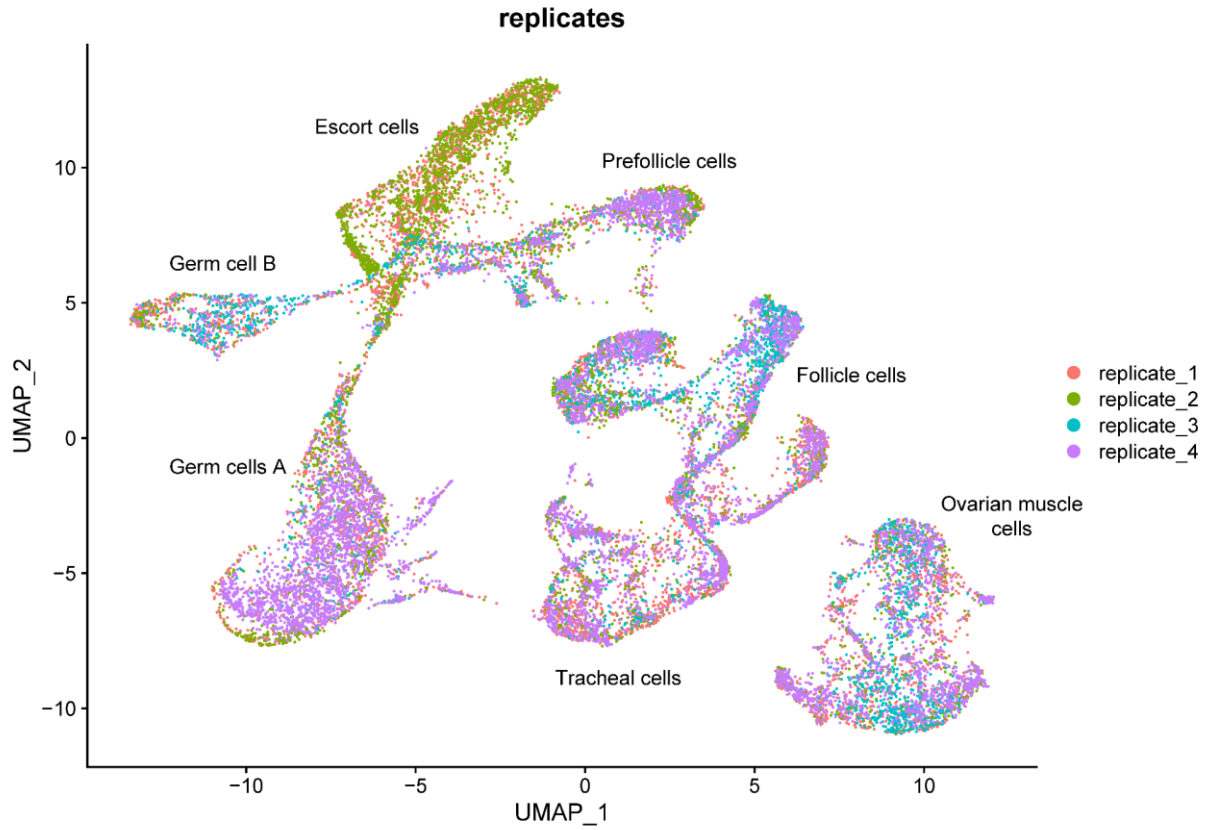
22. Turner JMA. https://doi.org/10.1146/annurev-genet-112414-055145. Annual Reviews; 2015 [cited 2024 Jan 30]. Meiotic Silencing in Mammals. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-112414-055145
23. X L, P M, Mj N, M S. Single-Cell RNA Sequencing of the *Cynomolgus Macaque* Testis Reveals Conserved Transcriptional Profiles during Mammalian Spermatogenesis. Dev Cell [Internet]. 2020 Aug 24 [cited 2024 Jan 29];54(4). Available from: https://pubmed.ncbi.nlm.nih.gov/32795394/
24. Handel MA. The XY body: a specialized meiotic chromatin domain. Exp Cell Res. 2004 May 15;296(1):57–63.
25. Turner JMA, Mahadevaiah SK, Ellis PJI, Mitchell MJ, Burgoyne PS. Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. Dev Cell. 2006 Apr;10(4):521–9.
26. Schoenmakers S, Wassenaar E, Hoogerbrugge JW, Laven JSE, Grootegoed JA, Baarends WM. Female Meiotic Sex Chromosome Inactivation in Chicken. Lee JT, editor. PLoS Genet. 2009 May 22;5(5):e1000466.
27. Guioli S, Lovell-Badge R, Turner JMA. Error-Prone ZW Pairing and No Evidence for Meiotic Sex Chromosome Inactivation in the Chicken Germ Line. Hassold TJ, editor. PLoS Genet. 2012 Mar 8;8(3):e1002560.
28. Strome S, Kelly WG, Ercan S, Lieb JD. Regulation of the X Chromosomes in *Caenorhabditis elegans*. Cold Spring Harb Perspect Biol. 2014 Mar 1;6(3):a018366–a018366.
29. Daish TJ, Casey AE, Grutzner F. Lack of sex chromosome specific meiotic silencing in platypus reveals origin of MSCI in therian mammals. BMC Biol. 2015 Dec;13(1):1–13.
30. Robben M, Ramesh B, Pau S, Meletis D, Luber J, Demuth J. scRNA-seq reveals novel genetic pathways and sex chromosome regulation in Tribolium spermatogenesis [Internet]. bioRxiv; 2023 [cited 2024 Mar 21]. p. 2023.07.18.549532. Available from: https://www.biorxiv.org/content/10.1101/2023.07.18.549532v1
31. Anderson J, Henikoff S, Ahmad K. Chromosome-specific maturation of the epigenome in the *Drosophila* male germline. eLife [Internet]. 2023 Nov 17 [cited 2024 Jan 29];12. Available from: https://elifesciences.org/reviewed-preprints/89373
32. Wei KHC, Chatla K, Bachtrog D. Single-cell RNA-seq of *Drosophila miranda* testis reveals the evolution and trajectory of germline sex chromosome regulation. PLOS Biol. 2024 Apr 30;22(4):e3002605.
33. Traut W, Schubert V, Daliková M, Marec F, Sahara K. Activity and inactivity of moth sex chromosomes in somatic and meiotic cells. Chromosoma. 2019 Dec;128(4):533–45.
34. Elkrewi M, Khauratovich U, Toups MA, Bett VK, Mrnjavac A, Macon A, et al. ZW sex-chromosome evolution and contagious parthenogenesis in Artemia brine shrimp. Dyer K, editor. Genetics. 2022 Sep 30;222(2):iyac123.
35. Jaglarz MK, Bilinski SM. Oogenesis in Crustaceans: Ultrastructural Aspects and Selected Regulating Factors. In: Cothran R, Thiel M, editors. Reproductive Biology: The Natural History of the Crustacea, Volume 6 [Internet]. Oxford University Press; 2020 [cited 2024 Feb 12]. p. 0. Available from: https://doi.org/10.1093/oso/9780190688554.003.0002
36. Abatzopoulos ThJ, Beardmore JA, Clegg JS, Sorgeloos P, editors. *Artemia*: Basic and Applied Biology [Internet]. Dordrecht: Springer Netherlands; 2002 [cited 2024 Feb 12]. Available from: http://link.springer.com/10.1007/978-94-017-0791-6
37. Xu LY, Wu WT, Bi N, Yan ZJ, Yang F, Yang WJ, et al. A cytological revisit on parthenogenetic *Artemia* and the deficiency of a meiosis-specific recombinase DMC1 in the possible transition from bisexuality to parthenogenesis. Chromosoma. 2023 Jun 1;132(2):89–103.
38. Schurko AM, Logsdon JM, Eads BD. Meiosis genes in *Daphnia pulex* and the role of parthenogenesis in genome evolution. BMC Evol Biol. 2009 Apr 21;9(1):78.
39. Gómez R, Van Damme K, Gosálvez J, Morán ES, Colbourne JK. Male meiosis in Crustacea: synapsis, recombination, epigenetics and fertility in *Daphnia magna*. Chromosoma. 2016 Sep;125(4):769–87.
40. Mikolajewicz N, Gacesa R, Aguilera-Uribe M, Brown KR, Moffat J, Han H. Multi-level cellular and functional annotation of single-cell transcriptomes using scPipeline. Commun Biol.

2022 Oct 28;5(1):1–14.
41. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2023 May 25;1–12.
42. Li H, Janssens J, Waegeneer MD, Kolluru SS, Davie K, Gardeux V, et al. Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. Science [Internet]. 2022 Mar 4 [cited 2024 Feb 8]; Available from: https://www.science.org/doi/10.1126/science.abk2432
43. Tarashansky AJ, Musser JM, Khariton M, Li P, Arendt D, Quake SR, et al. eLife. eLife Sciences Publications Limited; 2021 [cited 2024 Feb 5]. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. Available from: https://elifesciences.org/articles/66747
44. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008 Nov;9(9):1–9.
45. Duan H, Shao X, Liu W, Xiang J, Pan N, Wang X, et al. Spatio-temporal patterns of ovarian development and VgR gene silencing reduced fecundity in parthenogenetic *Artemia*. Open Biol [Internet]. 2023 Nov 15 [cited 2023 Nov 18]; Available from: https://royalsocietypublishing.org/doi/10.1098/rsob.230172
46. Lantz V, Chang JS, Horabin JI, Bopp D, Schedl P. The *Drosophila* orb RNA-binding protein is required for the formation of the egg chamber and establishment of polarity. Genes Dev. 1994 Mar 1;8(5):598–613.
47. Rust K, Byrnes LE, Yu KS, Park JS, Sneddon JB, Tward AD, et al. A single-cell atlas and lineage analysis of the adult *Drosophila* ovary. Nat Commun. 2020 Nov 6;11(1):1–17.
48. Hughes SE, Miller DE, Miller AL, Hawley RS. Female Meiosis: Synapsis, Recombination, and Segregation in *Drosophila melanogaster*. Genetics. 2018 Mar 1;208(3):875–908.
49. Christophorou N, Rubin T, Huynh JR. Synaptonemal Complex Components Promote Centromere Pairing in Pre-meiotic Germ Cells. PLOS Genet. 2013 Dec 19;9(12):e1004012.
50. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. Cell Rep Methods. 2023 Jun 26;3(6):100498.
51. Pang LY, DeLuca S, Zhu H, Urban JM, Spradling AC. Chromatin and gene expression changes during female *Drosophila* germline stem cell development illuminate the biology of highly potent stem cells. Bach EA, Banerjee U, editors. eLife. 2023 Oct 13;12:RP90509.
52. Bett VK, Macon A, Vicoso B, Elkrewi M. Chromosome-Level Assembly of *Artemia franciscana* Sheds Light on Sex Chromosome Differentiation. Genome Biol Evol [Internet]. 2024 Jan 5 [cited 2024 Feb 5];16(1). Available from: https://dx.doi.org/10.1093/gbe/evae006
53. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol. 2017 Jul 1;34(7):1812–9.
54. Wang J, Sun H, Jiang M, Li J, Zhang P, Chen H, et al. Tracing cell-type evolution by cross-species comparison of cell atlases. Cell Rep. 2021 Mar;34(9):108803.
55. Hayashi S, Kondo T. Development and Function of the *Drosophila* Tracheal System. Genetics. 2018 Jun;209(2):367–80.
56. Berg C, Sieber M, Sun J. Finishing the egg. Genetics [Internet]. 2024 Jan 3 [cited 2024 Jan 30];226(1). Available from: https://dx.doi.org/10.1093/genetics/iyad183
57. Kondo T, Hayashi S. Two-step regulation of trachealess ensures tight coupling of cell fate with morphogenesis in the *Drosophila* trachea. eLife [Internet]. 2019 [cited 2024 Jan 30];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6707767/
58. Mitchell B, Crews ST. Expression of the *Artemia* trachealess gene in the salt gland and epipod. Evol Dev. 2002 Sep 1;4(5):344–53.
59. Spradling AC, Niu W, Yin Q, Pathak M, Maurya B. eLife. eLife Sciences Publications Limited; 2022 [cited 2024 Jan 14]. Conservation of oocyte development in germline cysts from *Drosophila* to mouse. Available from: https://elifesciences.org/articles/83230
60. Timmons AK, Mondragon AA, Meehan TL, McCall K. Control of non-apoptotic nurse cell death by engulfment genes in *Drosophila*. Fly (Austin). 2017 Apr 3;11(2):104–11.
61. Yu Q, Liu X, Fang J, Wu H, Guo C, Zhang W, et al. Dynamics and regulation of mitotic

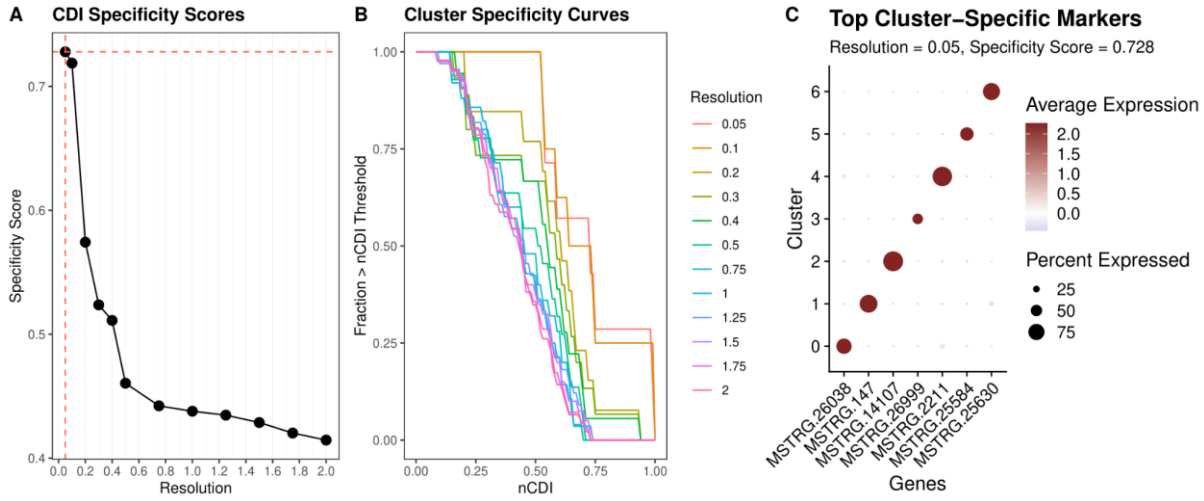
chromatin accessibility bookmarking at single-cell resolution. Sci Adv [Internet]. 2023 Jan [cited 2023 Dec 18]; Available from: https://www.science.org/doi/10.1126/sciadv.add2175
62. Ramos-Alonso L, Holland P, Le Gras S, Zhao X, Jost B, Bjørås M, et al. Mitotic chromosome condensation resets chromatin to safeguard transcriptional homeostasis during interphase. Proc Natl Acad Sci U S A. 120(4):e2210593120.
63. Munding EM, Shiue L, Katzman S, Donohue JP, Manuel Ares J. Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. Mol Cell. 2013 Aug 8;51(3):338.
64. Argyridou E, Parsch J. Regulation of the X Chromosome in the Germline and Soma of *Drosophila melanogaster* Males. Genes. 2018 May 4;9(5):242.
65. Chandra HS. How do heterogametic females survive without gene dosage compensation? J Genet. 1991 Dec 1;70(3):137–46.
66. Itoh Y, Melamed E, Yang X, Kampf K, Wang S, Yehya N, et al. Dosage compensation is less effective in birds than in mammals. J Biol. 2007 Mar 22;6(1):2.
67. Ellegren H, Hultin-Rosenberg L, Brunström B, Dencker L, Kultima K, Scholz B. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol. 2007 Sep 20;5(1):40.
68. Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC. Sex Chromosome-Specific Regulation in the *Drosophila* Male Germline But Little Evidence for Chromosomal Dosage Compensation or Meiotic Inactivation. PLOS Biol. 2011 Aug 16;9(8):e1001126.
69. Raz AA, Vida GS, Stern SR, Mahadevaraju S, Fingerhut JM, Viveiros JM, et al. Emergent dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of *Drosophila* testes. Buszczak M, Banerjee U, Buszczak M, Nystul TG, editors. eLife. 2023 Feb 16;12:e82201.
70. McKee BD, Yan R, Tsai JH. Meiosis in male *Drosophila*. Spermatogenesis. 2012 Jul 1;2(3):167–84.
71. Lukhtanov VA, Dincă V, Friberg M, Šíchová J, Olofsson M, Vila R, et al. Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. Proc Natl Acad Sci. 2018 Oct 9;115(41):E9610–9.
72. McLaughlin CN, Qi Y, Quake SR, Luo L, Li H. Isolation and RNA sequencing of single nuclei from *Drosophila* tissues. STAR Protoc. 2022 Jun 17;3(2):101417.
73. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nat Methods. 2020 Jun;17(6):615–20.
74. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019 Dec 16;20(1):278.
75. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8(1):14049.
76. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol. 2019 Aug;37(8):925–36.
77. Fleming SJ, Chaffin MD, Arduini A, Akkad AD, Banks E, Marioni JC, et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat Methods. 2023 Sep;20(9):1323–35.
78. Ranjan B, Sun W, Park J, Mishra K, Schmidt F, Xie R, et al. DUBStepR is a scalable correlation-based feature selection method for accurately clustering single-cell data. Nat Commun. 2021 Oct 6;12(1):1–12.
79. Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, dos Santos G, et al. FlyBase: a guided tour of highlighted features. Genetics. 2022 Apr 1;220(4):iyac035.
80. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018 Aug;560(7719):494–8.
81. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15.
82. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient

cell states through dynamical modeling. Nat Biotechnol. 2020 Dec;38(12):1408–14.
83. Speir ML, Bhaduri A, Markov NS, Moreno P, Nowakowski TJ, Papatheodorou I, et al. UCSC Cell Browser: visualize your single-cell data. Bioinformatics. 2021 Dec 7;37(23):4578–80.

Supplementary material

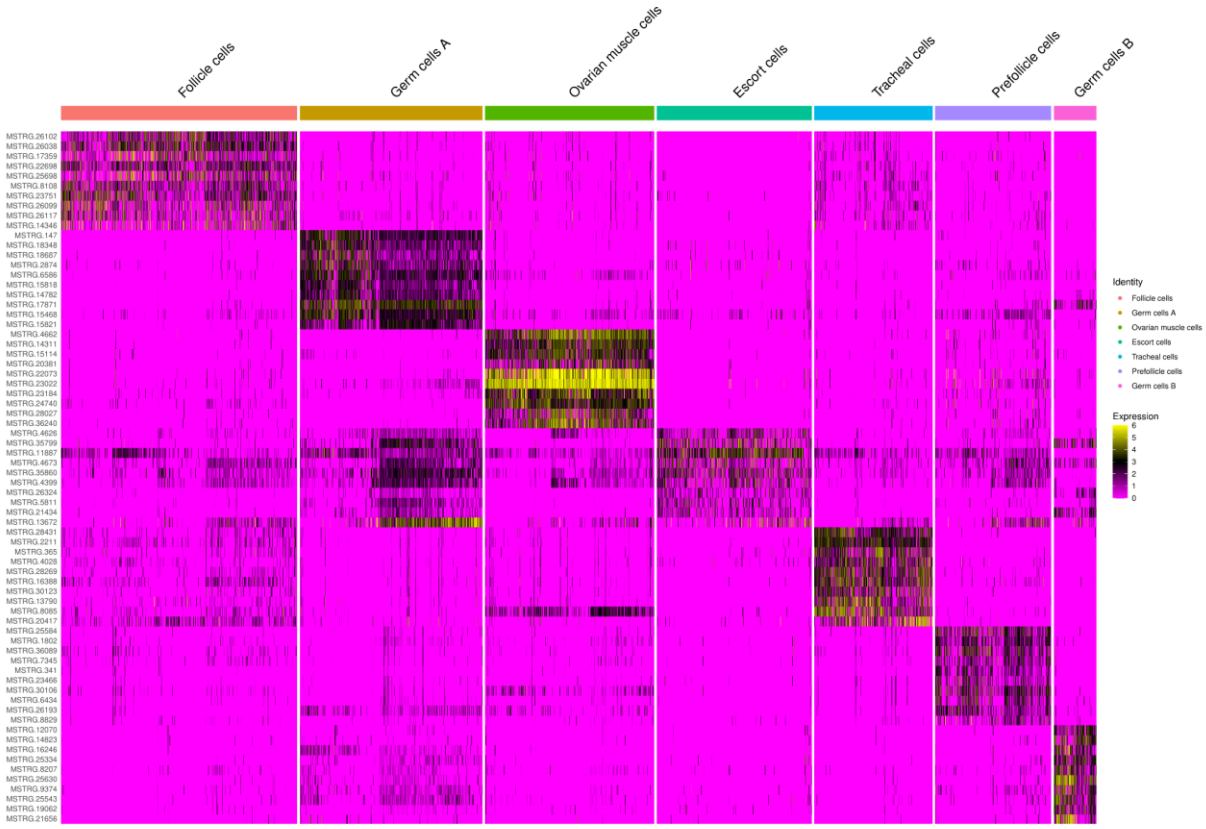


S1 Fig: UMAP of all nuclei coloured by replicate.

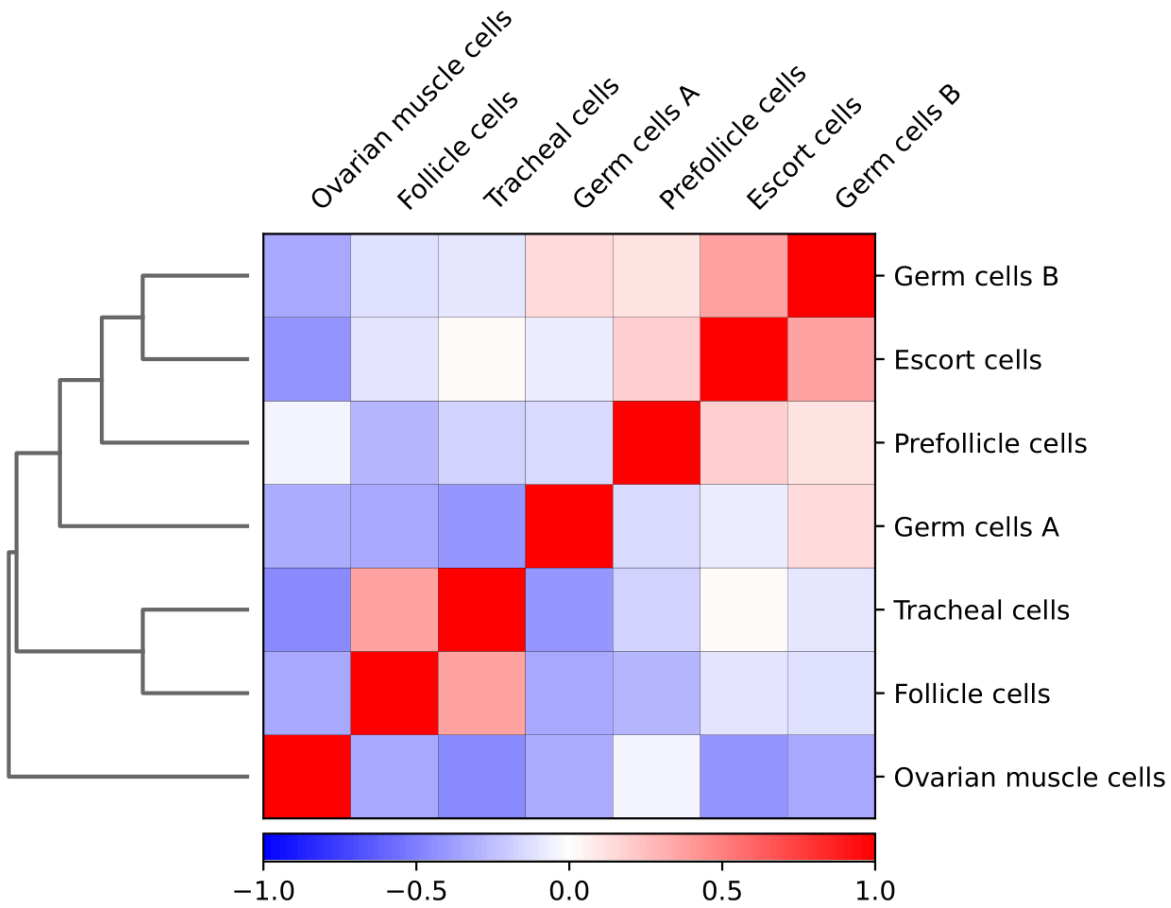


S2 Fig: A) specificity scores. B) Specificity curves. C) Dotplot of top cluster-specific markers.

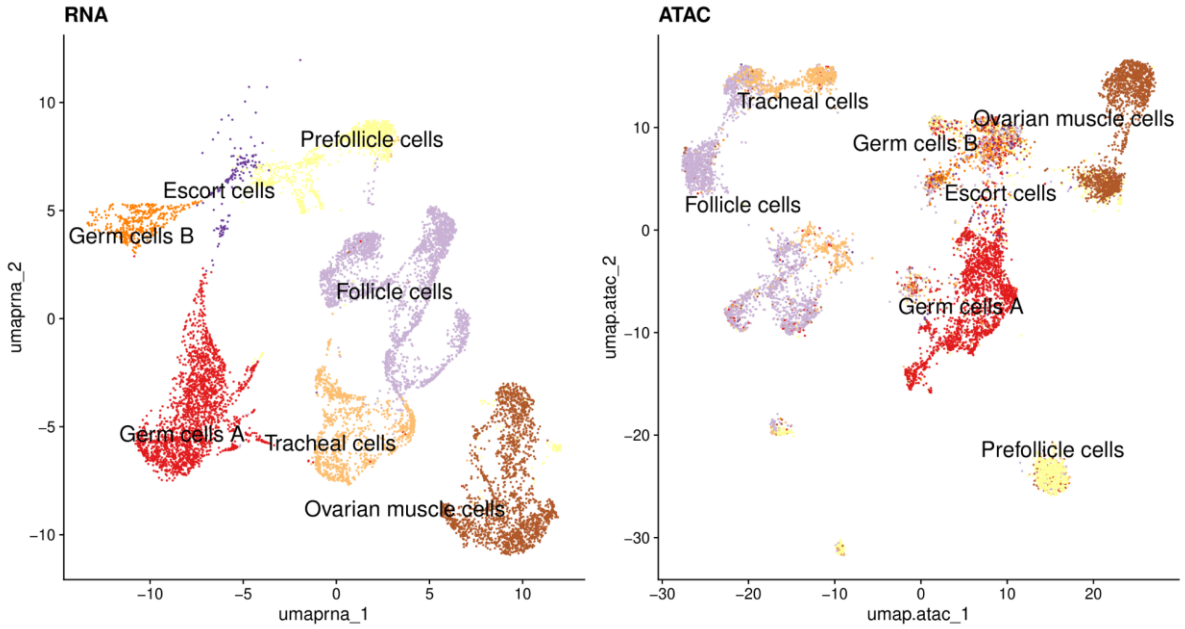
S3 Fig: Cluster specific markers based on findallmarkers.



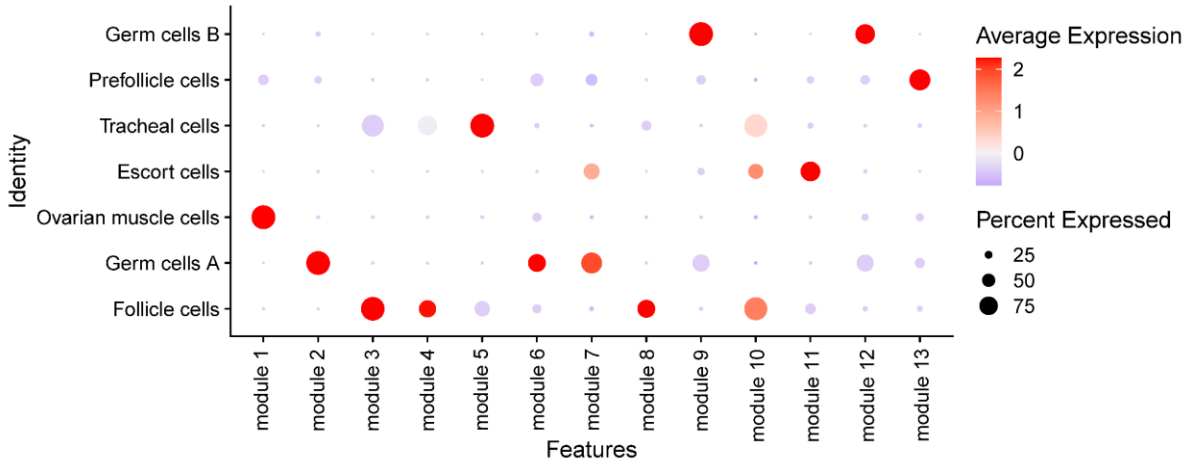
S4 Fig: Cluster specific markers based on findconservedmarkers.



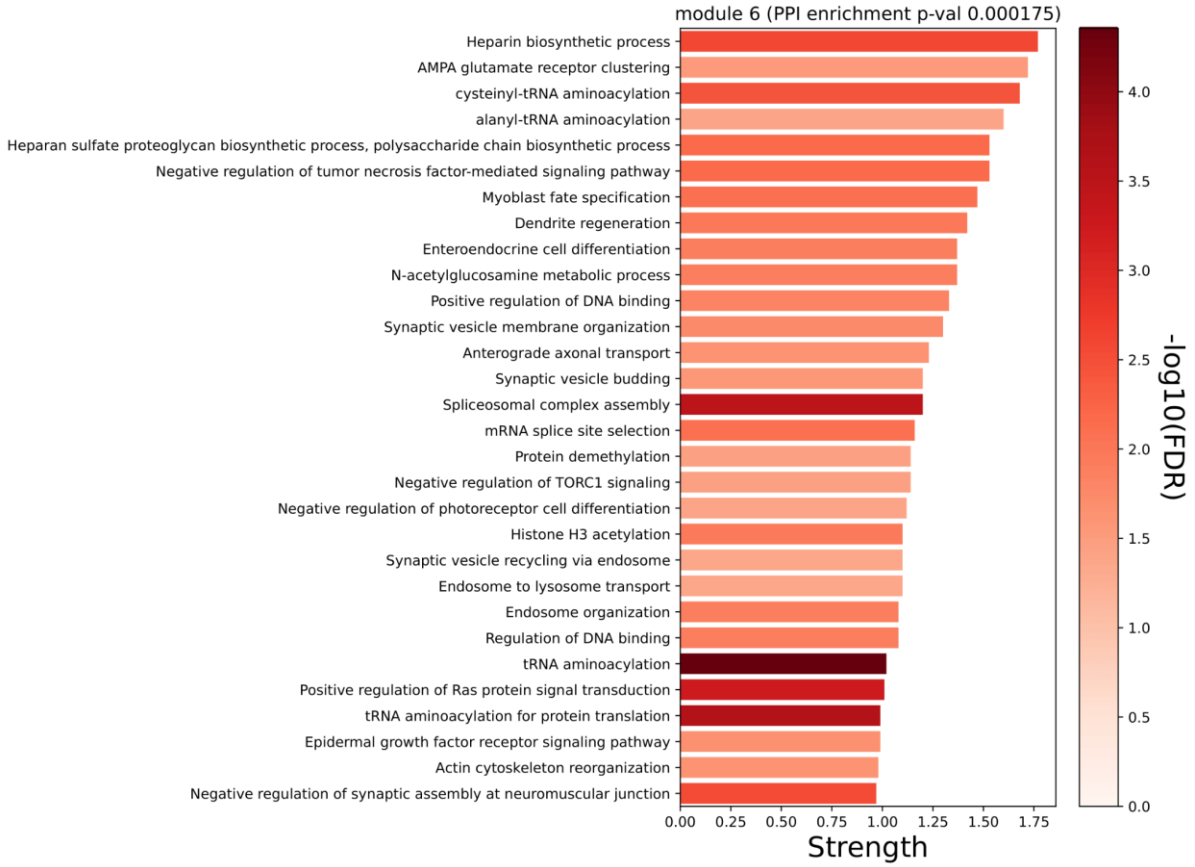
S5 Fig: A dendrogram and heatmap of the correlation matrix of the mean expression values per cluster (*Artemia*).



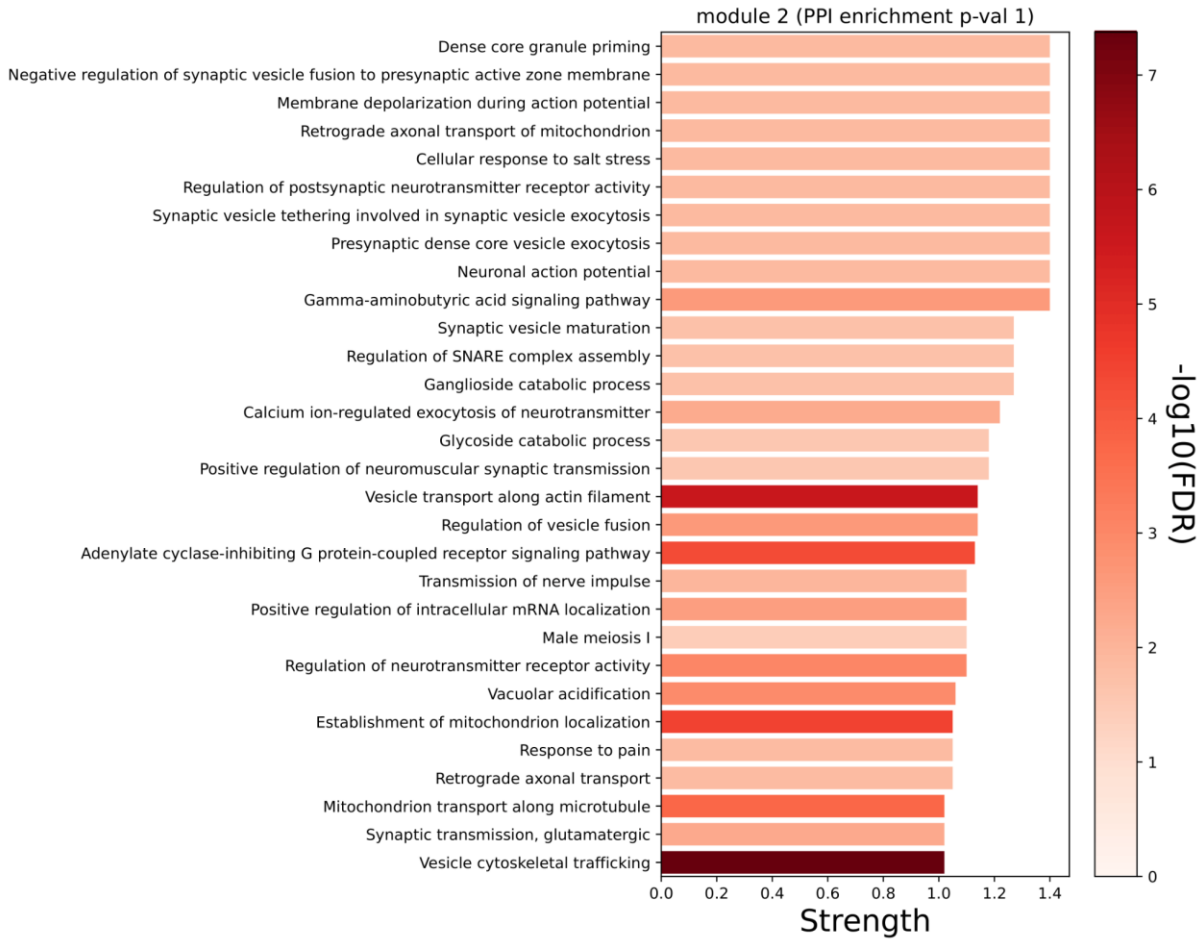
S6 FIg: A) UMAP of replicates 3 and 4 nuclei based on expression. B) UMAP of the nuclei from replicates 3 and 4 nuclei on peaks.



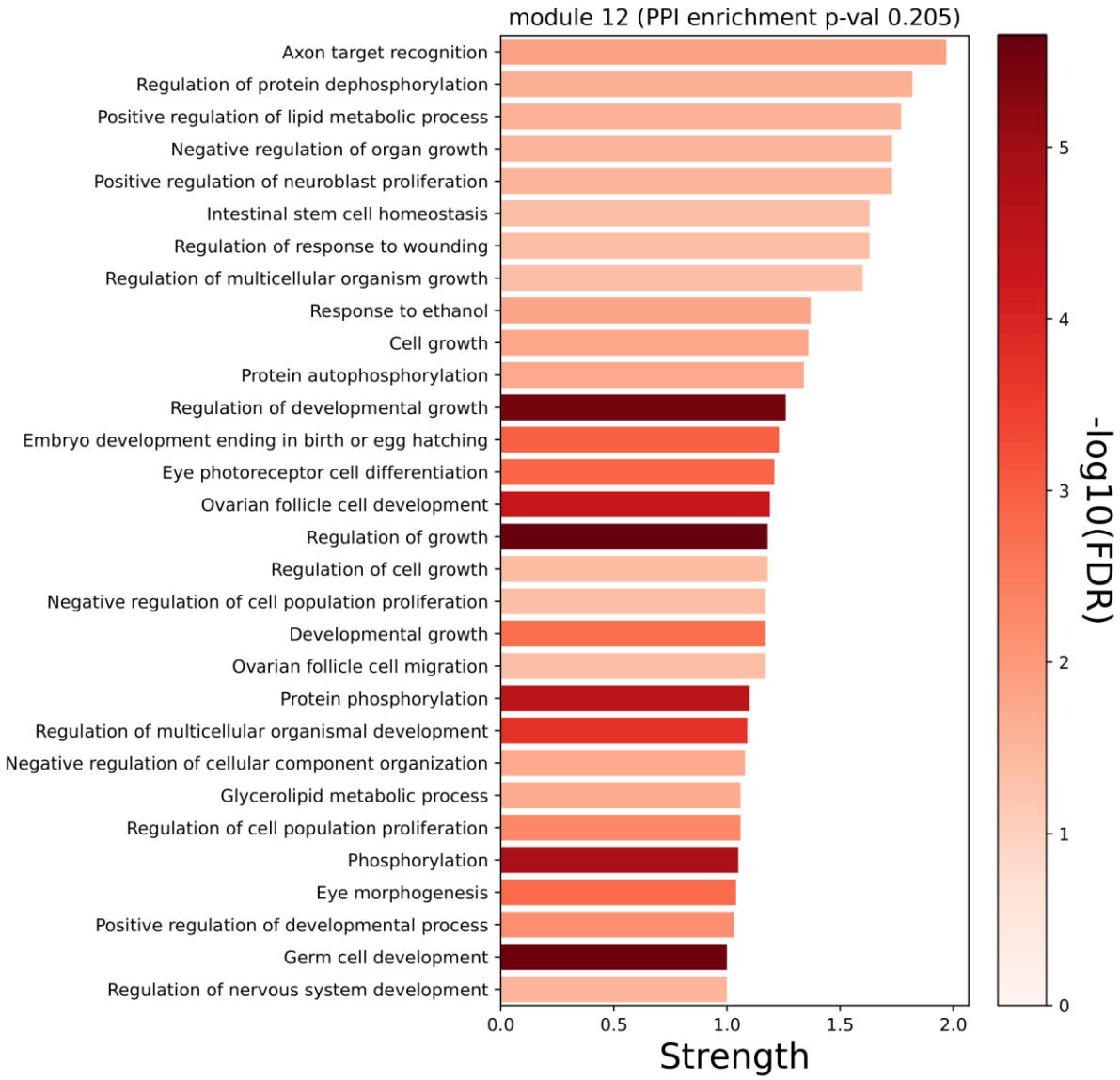
S7 Fig: Dotplot depicting the expression of the modules identified using the co-expression network analysis.



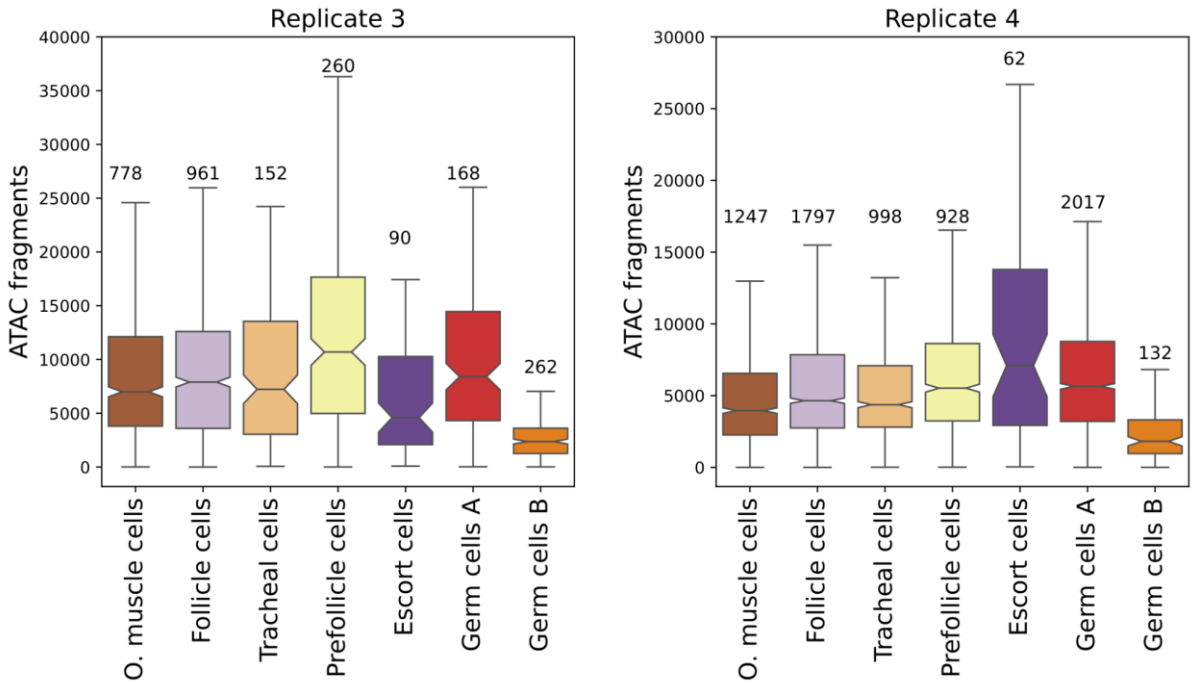
S8 Fig: Biological process GO enrichment in module 6 (433 genes).



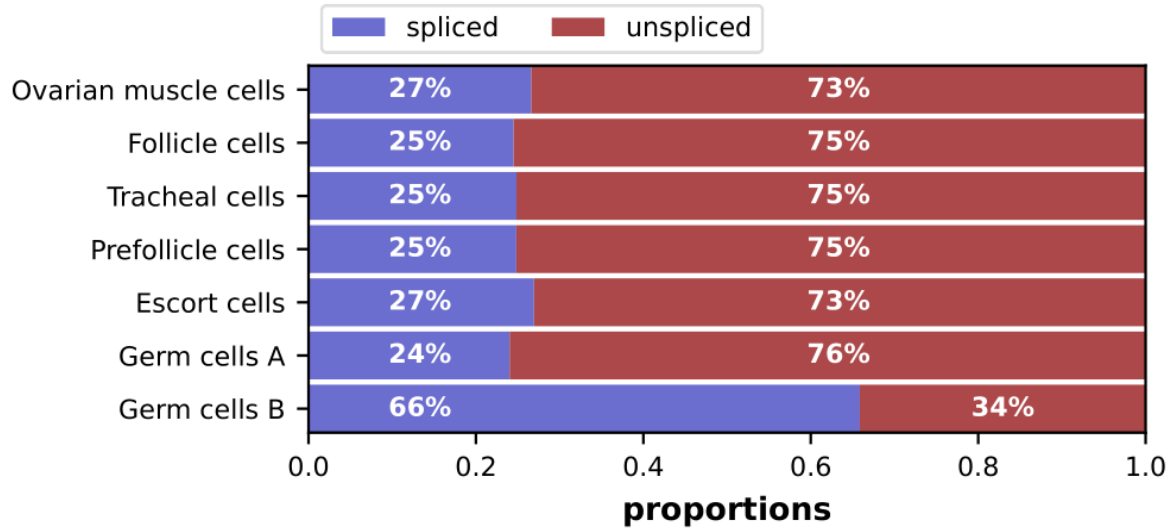
S9 Fig: Biological process GO enrichment in module 2 (1363 genes).



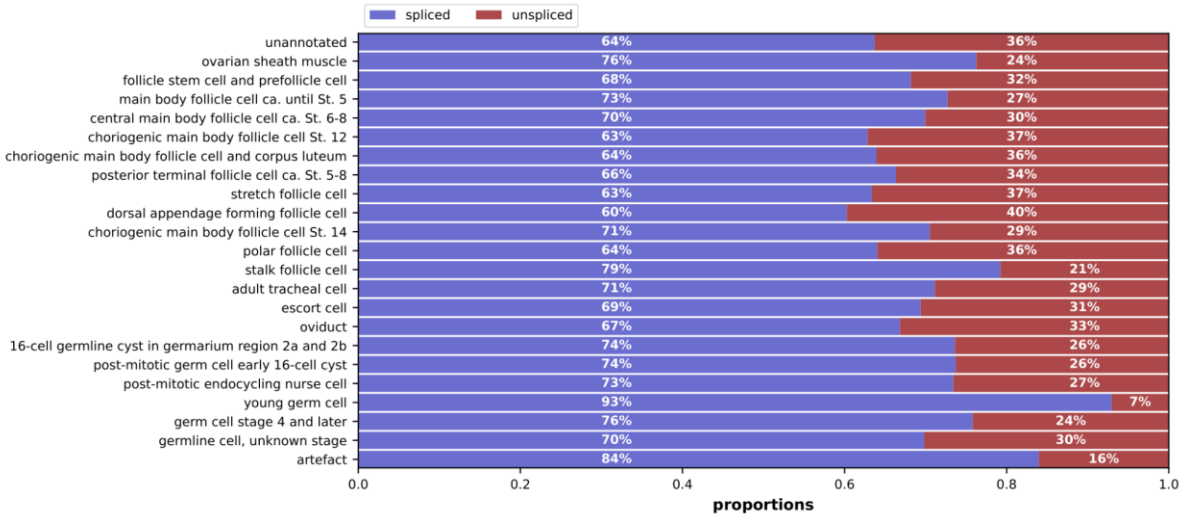
S10 Fig: Biological process GO enrichment in module 12 (61 genes).



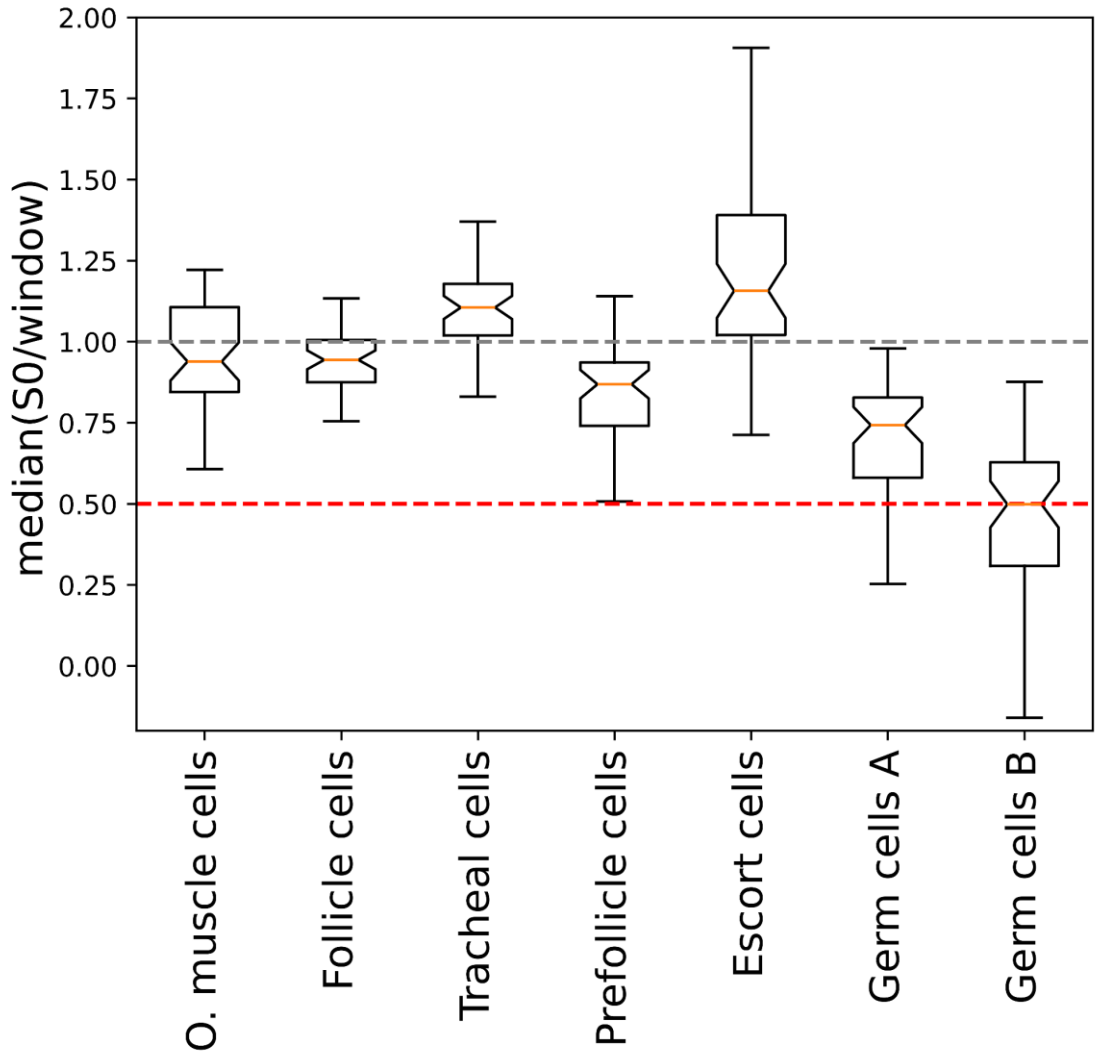
S11 Fig: A) ATAC fragments per cell for all clusters in Replicate 3. B) ATAC fragments per cell for all clusters in Replicate 4.



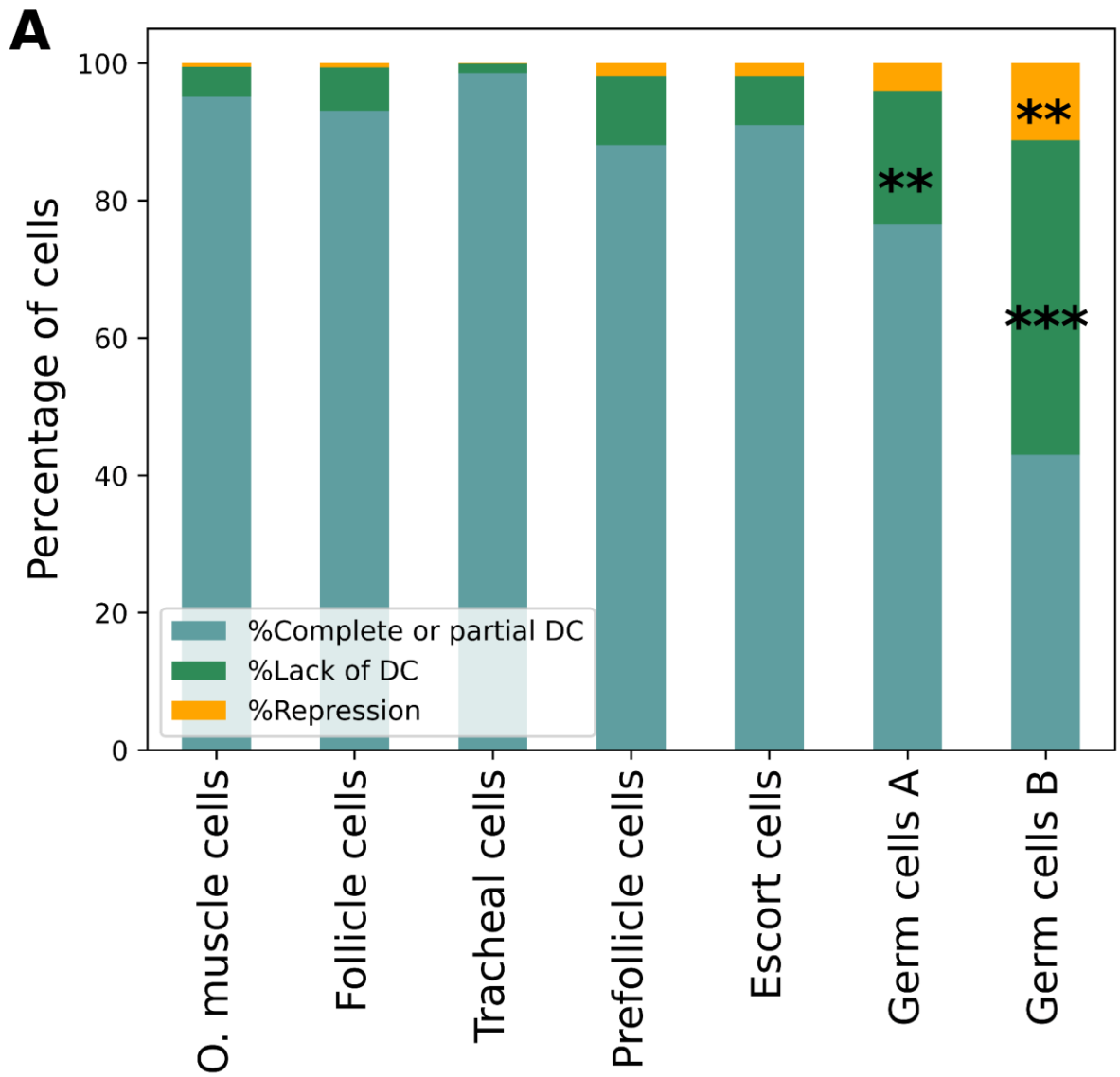
S12 Fig: The percentage of spliced and unspliced transcripts (replicates 1 and 2)



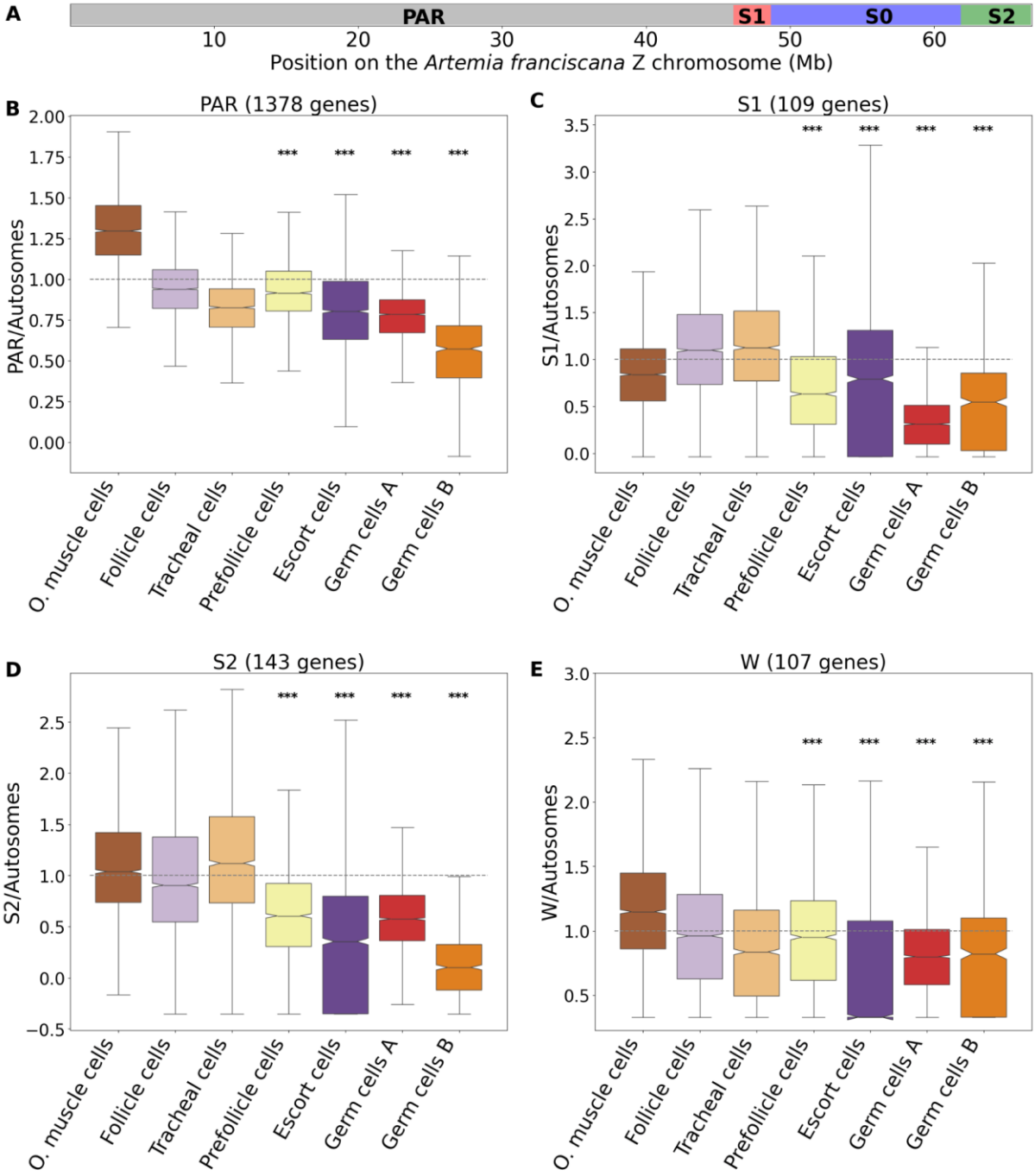
S13 Fig: *Drosophila* estimates of spliced and unspliced transcripts in the fly cell atlas data (raw data downloaded from NCBI bioproject PRJEB45570).



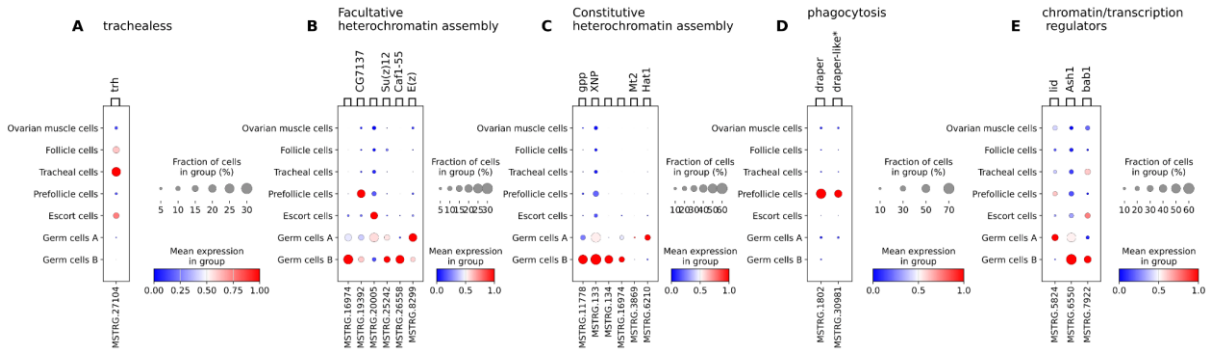
S14 Fig: Cluster medians of mean(S0)/mean(window) per cell) estimated using 49 non-overlapping genomic windows with the same number of genes as the S0.



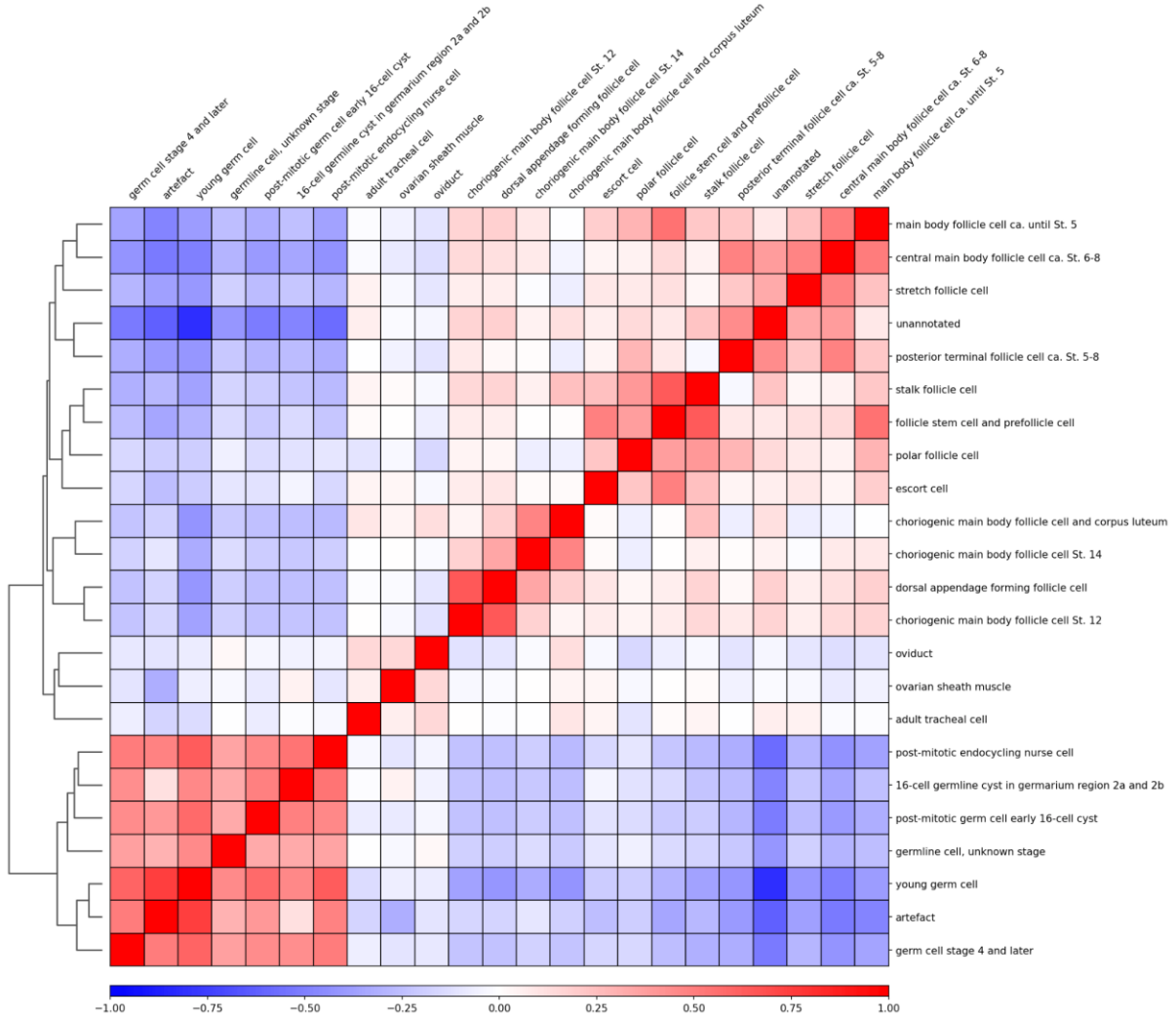
S15 Fig: The percentage of cells that are dosage compensated (DC), lack dosage compensation (Lack of DC) and repressed (Repression) using percentile-based cutoffs. The stars show the significance values comparing group 1 clusters and somatic clusters (%Lack dosage compensation vs rest, and %Repression vs rest using Chi-square contingency test). For the stars, *** denotes p-value <= 0.001, ** denotes p-value <= 0.01, and * denotes p-value <=0.05.



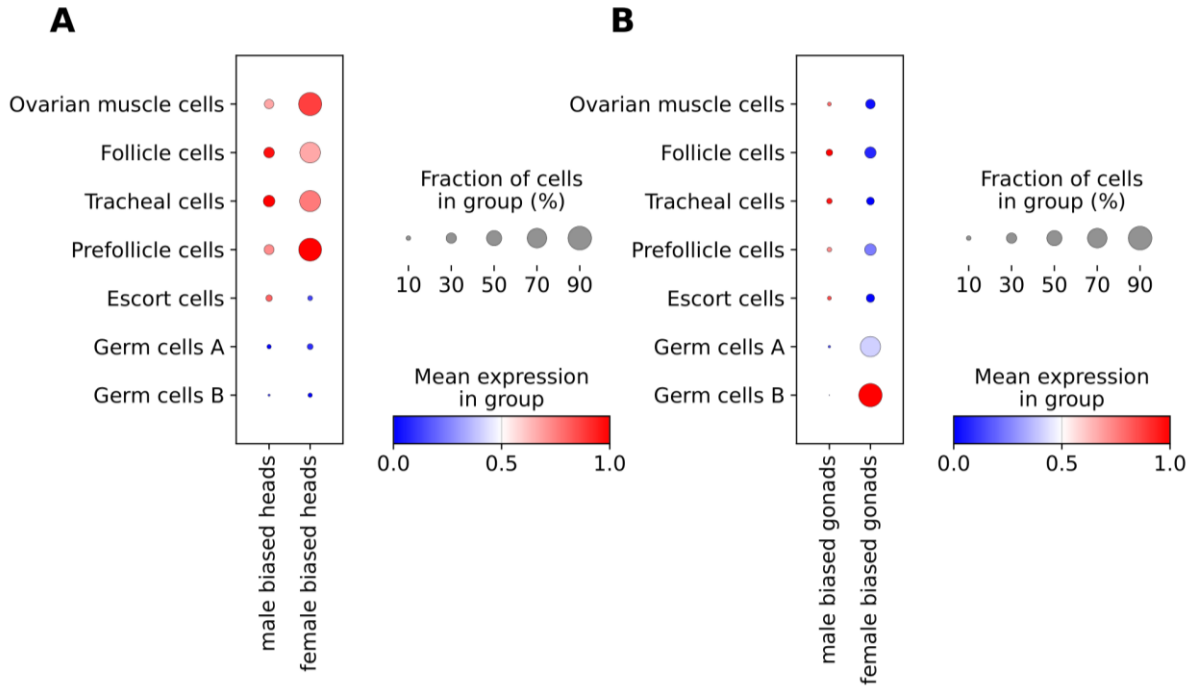
S16 Fig: A) The structure of the Z chromosome as described in (Bett et al., 2024), with the large pseudoautosomal region (PAR), the differentiation region (S0), and two younger strata (S1 and S2). B) PAR/Autosomes expression per cell C) S1/Autosomes expression per cell D) S2/Autosomes expression per cell E) W/Autosomes expression per cell. The normalized counts matrix was used for all the estimates.



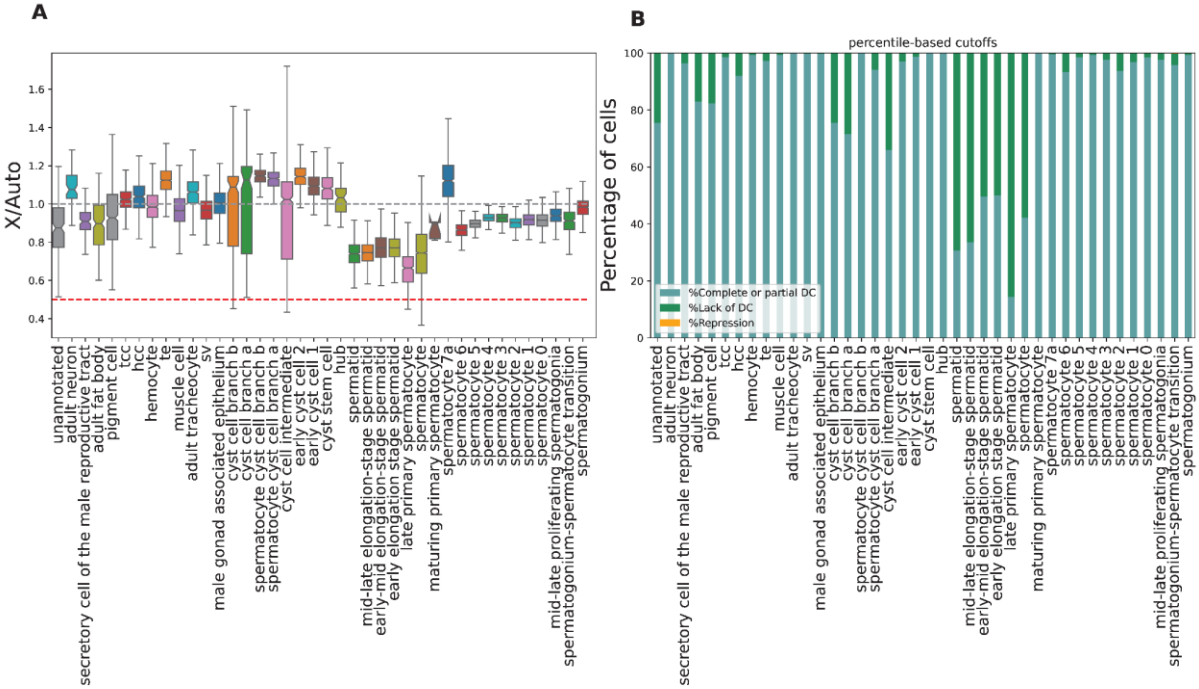
S17 Fig: A) Trachealess expression dotplot. B) Facultative heterochromatin assembly network expression. C) Constitutive heterochromatin assembly network expression D) phagocytosis genes expression. E) Genes involved in the modeling of oocyte chromatin.



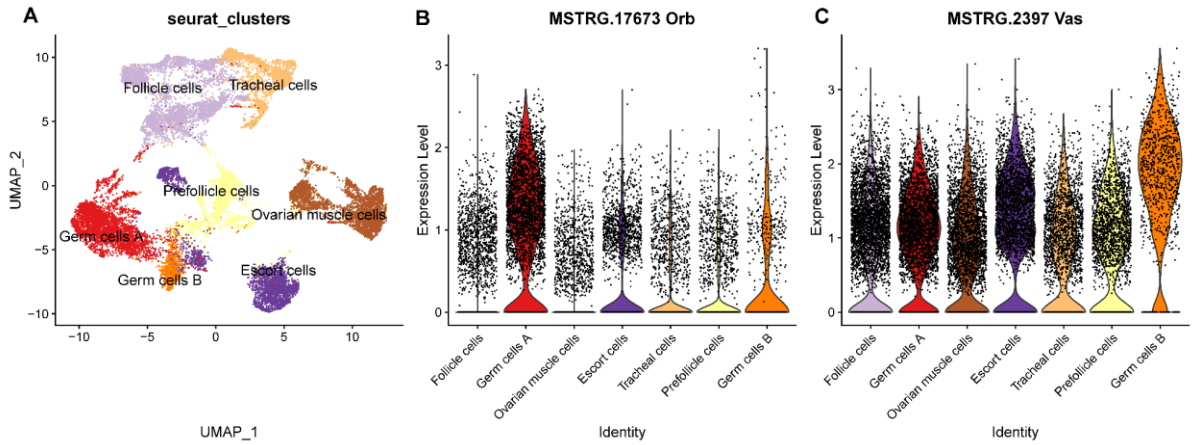
S18 Fig: A dendrogram and heatmap of the correlation matrix of the mean expression values per cluster in the fly cell atlas ovary data (10x, Stringent, H5AD, downloaded from https://flycellatlas.org/).



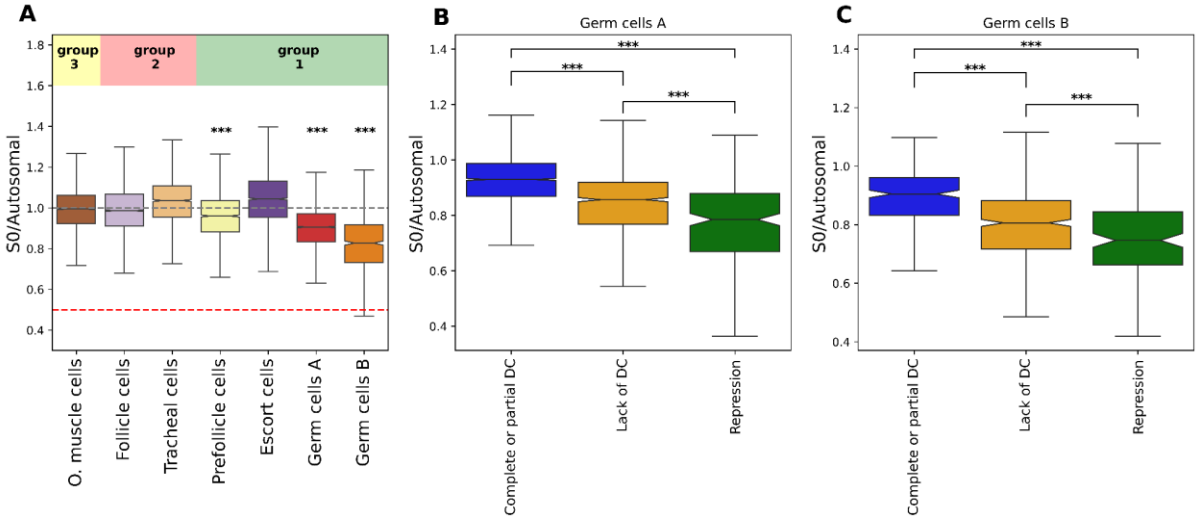
S19 Fig: Somatic clusters are enriched for genes that are female biased in heads (a somatic tissue), while germ cells B are enriched for female biased genes in the ovary. Male and female-biased genes were inferred by running DEseq2 with standard parameters on the bulk RNAseq data from (Huylmans et al., 2019). A) Differentially expressed male vs female heads (Filtered for >5,-5 fold change and <0.01 qval). B) Differentially expressed ovaries vs testes (Filtered for >10,-10 fold change and <0.01 qval).



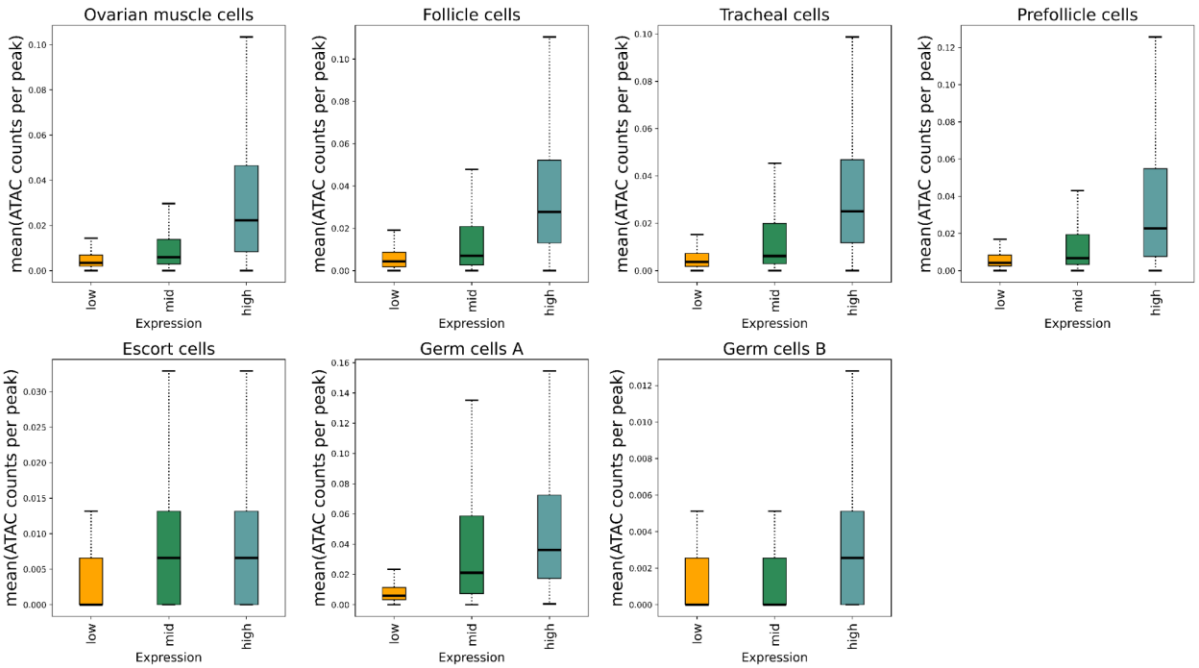
S20 Fig: Downregulation of the X chromosome in the testis from the *Drosophila* single nucleus atlas. A) The X/Autosomes expression per cell estimated using the normalized counts matrix. B) The percentage of cells that have partial or complete dosage compensation (Complete or partial DC), lack dosage compensation (Lack of DC), or repression using the percentile-based cutoffs. The testis snRNA-seq data (Raz AA et al., 2023) was obtained from: https://datadryad.org/stash/dataset/doi:10.5061/dryad.m63xsj454.



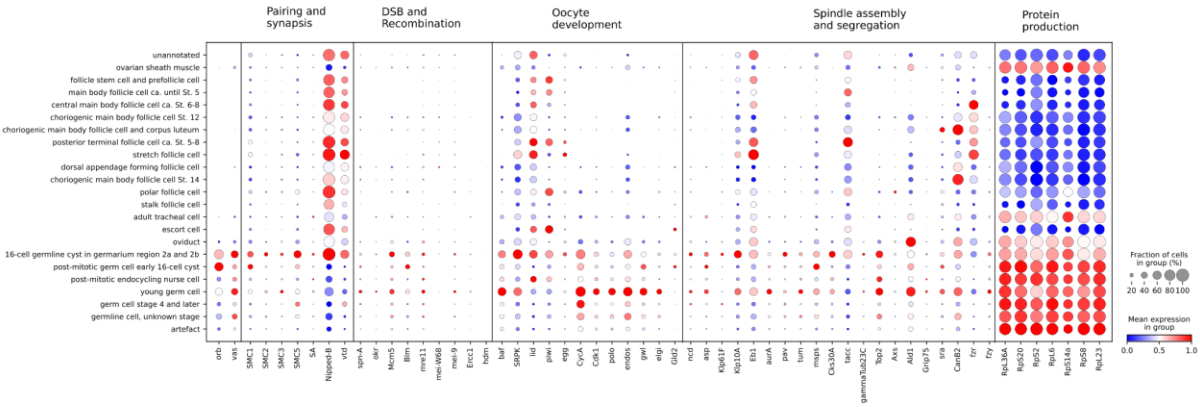
S21 Fig: A) UMAP with no ambient RNA removal. B) Orb expression no ambient RNA removal. C) Vasa expression with no ambient RNA removal. The cells were labeled based on the annotation from the main analysis.



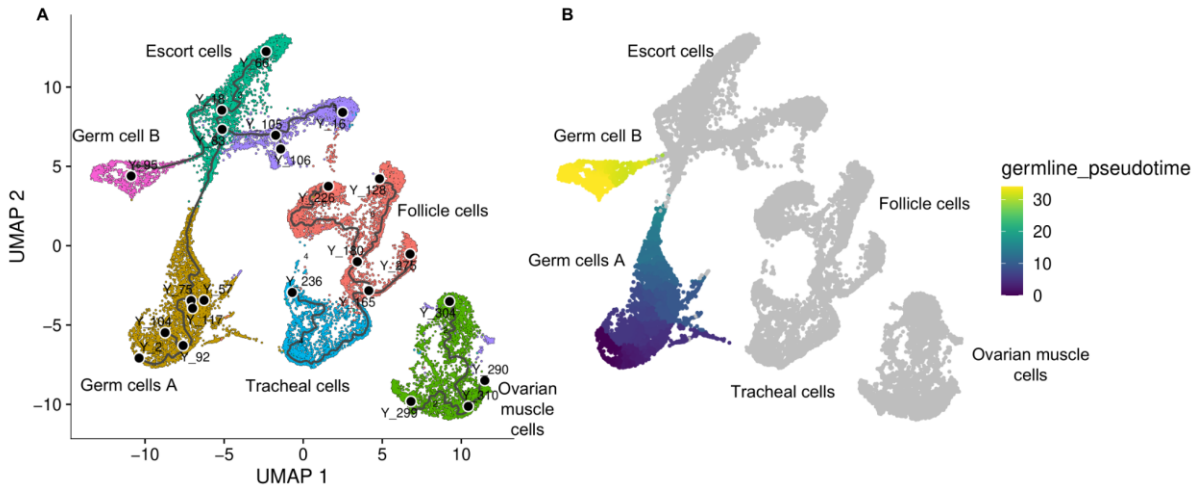
S22 Fig: A) S0/Autosomes expression per cell estimated using the normalized counts matrix (No ambient RNA removal) B) S0/Autosomal for the germ cells A cluster cells in the different expression zones (no ambient removal). C) S0/Autosomal for the germ cells B cluster cells in the different expression zones (no ambient removal).



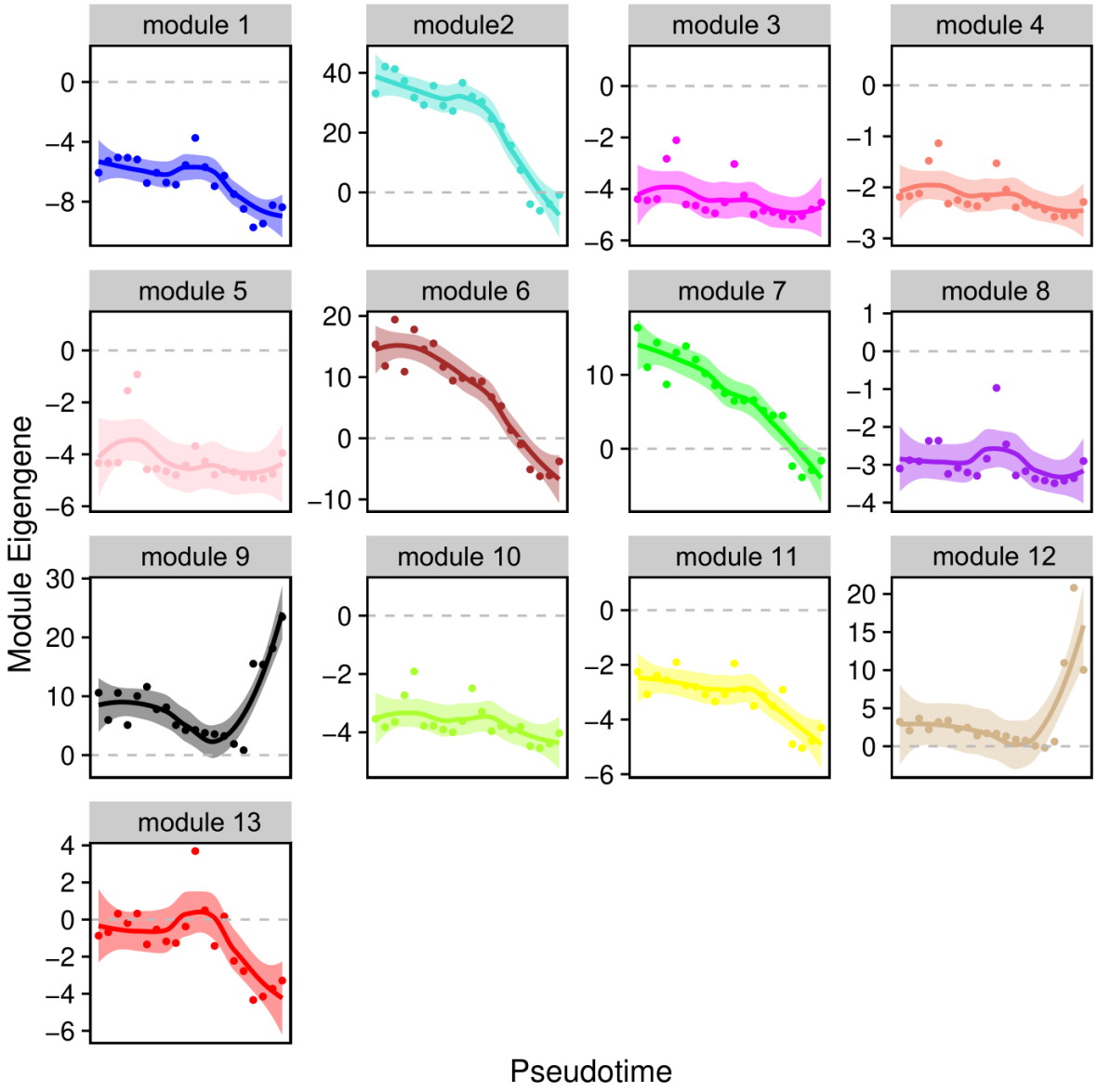
S23 Fig: Correlation between the mean expression of genes in each cluster and the mean of the ATAC counts in the linked peaks. Genes were split into three categories: <20% for low expression genes, >20% and <80% for medium expression, and high expression genes >80%.



S24 Fig: The expression of vas and orb, and the genes involved in the different stages of meiosis in *Drosophila*, along with the expression of genes involved in protein production. The plot was produced with Scanpy using the Fly Cell Atlas ovary dataset (10x, Stringent, H5AD, downloaded from https://flycellatlas.org/).



S25 Fig: A) UMAP depicting the Monocle3 pseudotime trajectory of all the clusters (all replicates). B) germline pseudotime (Germ cells A and Germ cells B). Y_2 (panel A) was used as the principal node.



S26 Fig: Module expression dynamics across the germline pseudotime.

S1 Table: Number of cells in the Expression UMAP from the females allowed to mate and the females not allowed to mate.

	#cells		
cluster	Replicates 1 and 2 (Mated)	Replicates 3 and 4 (Unmated)	(mated/(mated+unmated))*100
Ovarian muscle cells	1305	2032	39.10%
Follicle cells	1890	2768	40.60%
Tracheal cells	1175	1153	50.50%
Prefollicle cells	1097	1193	47.90%
Escort cells	2903	152	95%
Germ cells A	1402	2199	38.90%
Germ cells B	445	395	53%

S2 Table: Ambient RNA percentage estimated using Souporcell.

	%ambient RNA
Replicate 1	47.57
Replicate 2	45.26
Replicate 3	58.25
Replicate 4	39.63

CHAPTER 5

A candidate Z-linked gene for Asexual Reproduction in *Artemia* Brine Shrimp

Chapter 5: A candidate Z-linked gene for Asexual Reproduction in *Artemia* Brine Shrimp

Authors:
Marwan Elkrewi[1]@, Dominik Kopcak[2], Ariana Macon[1], Beatriz Vicoso[1]@
@corresponding authors
[1] Institute of Science and Technology Austria (ISTA), Klosterneuburg 3400, Austria
[2] Okinawa Institute of Science and Technology (OIST), 1919-1 Tancha, Onna-son, Kunigami-gun
Okinawa, Japan
Key words:
Meiosis, Asexual reproduction, parthenogenesis, single-nucleus RNAseq

Abstract
Although most animals reproduce sexually, transitions from sexual to asexual reproduction are well-documented across different taxa. Invertebrates seem to have a higher occurrence rate, which makes them suitable for studying the basis and consequences of asexual reproduction. Despite extensive efforts, the regulatory changes that lead to the emergence of asexuality remain largely undiscovered in the majority of species studied. *Artemia* brine shrimp are among the species where a long-standing interest exists in identifying the genetic basis of parthenogenesis; the presence of multiple sexual and obligate parthenogenetic species makes them a perfect model for this endeavor. The efforts so far have resulted in a model of the modified form meiosis in the species, a likely role for the Z-chromosome in its transmission, and a set of differentially expressed genes between multiple sexual and asexual species. However, no master regulator or genetic changes have been put forward as the root causes for the shift. In this paper, we tackle the question of asexuality using newly generated single nucleus RNAseq data of the female reproductive system of individuals from the Aibi lake population of *Artemia parthenogenetica* and the closely related species *Artemia sp. Kazakhstan*. We identify the germline cell clusters in the female reproductive system and perform differential expression and splicing analyses to identify the transcriptional differences between the meiotic cells of the two species. In combination with whole genome sequencing (WGS) data from a new backcrossing experiment and the published WGS data, we were able to identify a promising candidate for asexuality on the sex chromosomes; homologs of this gene are known to be important for various aspects of oocyte development.

Introduction

Many mysteries surround the evolution of reproduction and the varying success of its different forms. At the core of sexual reproduction is meiosis, which seems to have originated in early Eukaryotes more than 1 billion years ago (1). A cell that undergoes meiosis produces daughter cells with half the number of chromosomes and new allele combinations. This is achieved by doubling the number of chromosomes, pairing the homologous ones together and mediating recombination events between them, and then dividing twice to create the two daughter cells (2). The resulting reshuffling of genes and the creation of new allelic combinations is thought to be the reason behind the success of sexual reproduction (3).

The vast majority of animals and plants reproduce sexually (4), with germ cells from both parents undergoing meiosis to produce haploid gametes that fuse to restore diploidy and give rise to offspring that are a genetic mixture of their parents. On the other hand, many other organisms reproduce asexually, with some even alternating between a few rounds of asexual reproduction and a round of sexual reproduction, such as Aphids (5). Asexual reproduction can take place through various mechanisms that can be divided into two major forms with different evolutionary repercussions: clonal or ameiotic reproduction (apomixis) and modified meiosis (automixis). In the clonal case, which is the mode of reproduction of bacteria and many other unicellular and some multicellular organisms, the offspring are largely identical

genetic replicas of the parent (5). On the other hand, automixis, which takes place in females, can produce offspring that are distinct from or identical to their mother depending on the recombination levels and the form it takes. For instance, in central fusion automixis, meiosis proceeds normally, and then the egg fuses with the polar body from Meiosis I to restore diploidy. In this scenario, maternal heterozygosity is maintained only when no recombination events occur (3,6). However, when recombination events take place, they lead to the loss of heterozygosity in some loci, depending on their proximity to the centromere, with centromere-distal regions more likely to lose heterozygosity. Central fusion automixis has been reported in hymenopterans, dipterans, moths, crustaceans, and nematodes (5).

For a while, there was an intense debate on the mode of reproduction of the diploid parthenogenetic lineages of *Artemia*, as classical cytogenetics studies were inconsistent. Recent evidence from a population genetics study of *Artemia parthenogenetica* from two western Mediterranean populations suggested that the changes in heterozygosity patterns over two years are consistent with automixis by central fusion and low but nonzero recombination rates (6). A more recent paper compared the meiotic progression of *Artemia franciscana* and a parthenogenetic species using a combination of cytology, immunofluorescence and RNAseq. They observe that the parthenogenetic oocytes do not arrest at metaphase I and that after meiosis I, they do not divide and meiosis II is replaced by a mitosis-like division (7). In terms of molecular pathways, they found that Dmc1 and Rad51, known to be involved in homologous recombination, are upregulated in the sexual species, and knocking down Dmc1 results in issues during homologous pairing (7).

Despite the great efforts, to date, there is no strong evidence pointing towards the genetic changes that resulted in the shift to asexuality in the species and the downstream genes are affected by the changes. An earlier study from our lab compared various sexual and asexual species in the Eurasian clade of *Artemia* in somatic and gonadal tissues and identified several differentially expressed genes between the different reproductive modes using bulk RNAseq (8). The analysis did not identify clear differences, potentially due to the averaging that takes place as a result of pooling large numbers of cells together and the fact that meiotic cells represent a small minority of cells in the tissue. To circumvent those challenges, we use single-nucleus RNA sequencing on the female reproductive system asexual (Aibi Lake population) and sexual (*Artemia sp. kazakhstan*) females to be able to look at the transcriptomes of the individual cells and compare gene expression between the meiotic cells of the sexual and asexual species. The single cell resolution allows us to capture cells at different time points in the meiotic progression and identify the differences between the sexual and asexual transcriptional pathways. Additionally, the parthenogenetic species of *Artemia* are able to produce functional males at very low frequencies (9) through rare recombination events between the Z and W chromosome (10,11). Such “rare males” can mate with closely related sexual females to produce new asexual lineages, a phenomenon referred to as contagious parthenogenesis. We utilized the occurrence of three rare males from a different lineage of *Artemia parthenogenetica* (Atanasovsko population) to generate backcrosses to a sexual species (*Artemia sp. Kazakhstan*). The F2 females were then used to check if the patterns of inheritance support the presence of an asexuality locus on a region of the Z chromosome identified previously using a similar experiment with a rare male from Aibi Lake, (10). Finally, we combined the two approaches to detect promising candidate genes for the evolution of asexuality.

Results

Backcrossing experiment provides further evidence for a Z–linked asexuality locus

Previous results from our lab (10) have pointed to a role for the Z chromosome in the transmission of asexuality using a backcross of an *Artemia parthenogenetica* (Aibi lake) rare male and a sexual *Artemia sp. Kazakhstan* females. We have repeated the experiment using a cross between three rare males from *Artemia parthenogenetica* (Atanasovsko) and *Artemia sp. Kazakhstan* females. The offspring from the F1 were isolated into individual vials and F1 males were mated with unmated *Artemia sp. Kazakhstan* females. From the F2s, we froze and sequenced 9 females which reproduced asexually and 9 control females (frozen after 3

months of isolation without reproducing any offspring). We aligned the WGS sequencing data from the two experiments to the *Artemia sp. Kazakhstan* genome assembled using short reads and scaffolded using the *Artemia sinica* chromosome-level genome assembly. We called SNPs and estimated Fst for all biallelic SNPs. Both experiments show a region of elevated Fst on the Z chromosome (Chromosome 1). We used local PCA on windows of 20,000 SNPs to further characterize the patterns of relatedness/inheritance between the asexual and control females across the Z chromosome, and we see that there is strong separation by reproductive mode in both experiments (between windows 50 and 80). No other genomics regions show consistent patterns across the two experiments.

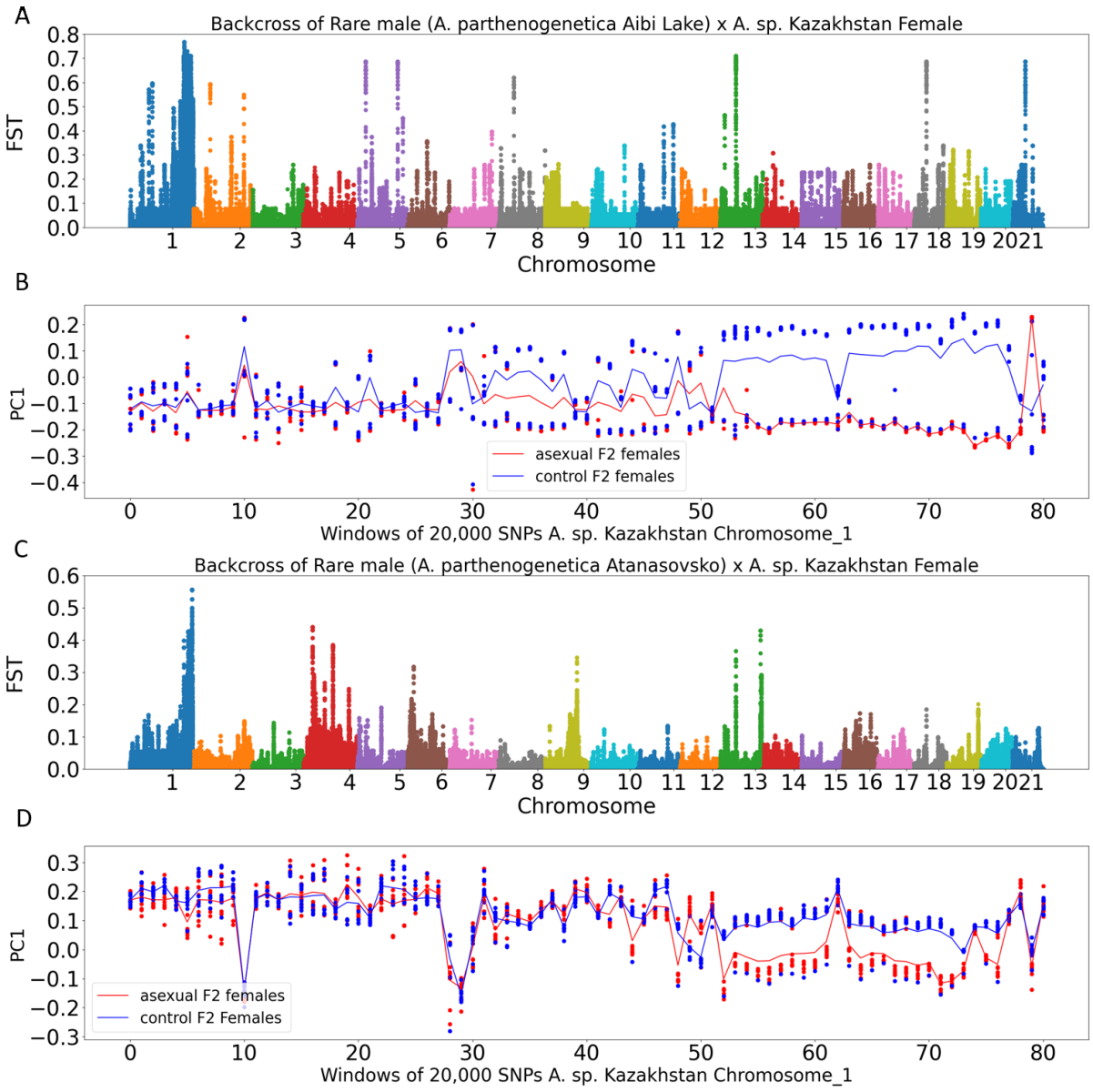


Figure 1: Backcrossing experiments suggest a role for the Z chromosome in the transmission of asexuality. A) Fst between the asexual and control females from the A. parthenogenetica (Aibi lake) cross across the different chromosomes. B) Local PCA analysis on windows of 20,000 SNPs across the Z chromosome between the asexual and control females (Aibi backross). B) Fst between the asexual and control females F2s from the A. parthenogenetica (Atanasovsko) backcross across the different chromosomes. B) Local PCA analysis on windows of 20,000 SNPs across the Z chromosome between the asexual and control females (Atanasovsko backross).

snRNAseq provides insights into the (mis)regulation of meiotic programs in the asexual germline cells

In order to identify the genes involved in the switch to asexual meiosis, we performed single-nucleus RNA sequencing on two replicates (comprising 25 females each) of the female reproductive system of *Artemia sp. Kazakhstan* and two replicates of *Artemia parthenogenetica* (Aibi lake) with the same number of females. We performed the analysis twice: once with the A. sp. Kazakhstan genome and once with the *A. parthenogenetica* genome, which we assembled from genomic short reads and scaffolded using the Artemia sinica genome (10) (see section Genome Assemblies and Annotation). This was to ensure that we do not miss any candidates which might be missing from either genomes and that the general patterns we recover are not dependent on the genome used. We used the same pipeline described in (12) to cluster the cells, and we annotated the different cell types using SAMAP by mapping to the *Artemia franciscana* atlas (Figure 2 C) (12). We merged the clusters that mapped to a single cell type in *Artemia franciscana* (Clusters 0 and 7 mapped to Escort cells, and clusters 5 and 6 mapped to Prefollicle cells, Supplementary Table 1), but we did not split any clusters (*A. franciscana* Tracheal and Follicle cells were recovered as a single cluster in the two analyses performed here). For the germline clusters, we confirmed the expression of the germline markers Orb and Vas, and the enrichment of homologs of the *Drosophila* meiosis-related genes identified in (12) (Figure 2 D,E, and F). Most of the cells retained the same identity in the two analyses, with the exception of Prefollicle cells, of which a large subset was assigned as Escort cells in analysis with the *A. parthenogenetica* reference (Supplementary Figures 1,2, 3 and 4).

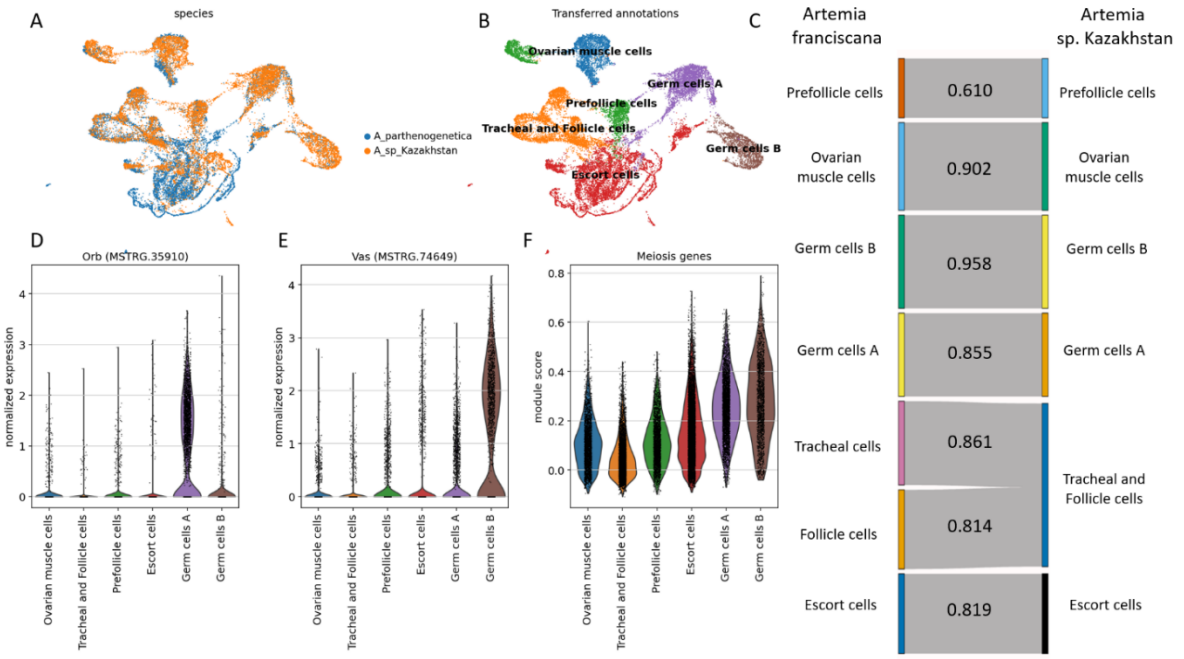


Figure 2: snRNAseq identified the different cell types in the female reproductive system of the two *Artemia* species. A) The UMAP plot of the different cells colored by the different species. B) UMAP plot of the different clusters annotated using the *Artemia franciscana* atlas. C) A sankey diagram showing the correspondence and alignment rate of the different clusters between the atlas generated in this study and the *Artemia franciscana* atlas. D) The expression of the early germline marker (Orb) in the different clusters. F) The expression of the late germline marker (Vasa) in the different clusters. F) Module score for the meiosis genes used in (12) in the different clusters.

In order to identify genes that could be involved in the regulation of asexual meiosis, we performed differential expression analysis between the two species within each cluster. To minimize the number of false positives, we used two complementary approaches. The first was to aggregate the expression counts within each cluster, which results in two replicates per species, and then using DESeq2 to test for differential expression (volcano plot of Germ

cells A is shown in Figure 3 B). The second approach used the Seurat FindMarkers function to find the differentially expressed genes between all the cells of one cluster in one species against the other, with MAST for differential testing and the replicates as a latent variable (volcano plot of Germ cells A is shown in Figure 3 A). We selected the genes that were significantly biased in the two analyses (p-adj <0.01 and abs(log2(fold change)) > 0.5) (Figure 3 C). The germ cells A cluster had the largest number of DE genes (330), followed by Prefollicle cells (223) (Figure 3 D). Prefollicle cells showed some similarity to germline cells in the *Artemia franciscana* paper (12), and they had the lowest mapping rate to the *Drosophila* clusters (0.230). Surprisingly, the Germ cells B cluster only had 28 DE genes, which could be due to the small number of cells recovered in the asexual species (1535 in Kaz vs 351 in Aibi) or to a lack of biological differences at that stage of oogenesis. As germ cells A showed the most significant differences between the two species, we found it likely that the transcriptional changes that lead to asexual meiosis might be happening at that stage of germline development. To functionally annotate the differentially expressed genes detected at this stage, we uploaded a transcriptome of 28,707 sequences to TRAPID, curated by selecting the isoforms with the longest open reading frame (ORF) for each gene and excluding transcripts with ORFs < 100 aa. Out of the 330 DE genes in Germ cells A, 192 were represented in the transcriptome (the others were not included due to the absence of a long ORF). We ran the GO enrichment analysis using the 192 transcript and then realized 4 transcripts were fragments of the same gene (inositol 1,4,5-trisphosphate receptor-like), which is a large protein in *Artemia franciscana* (~2600 aa, NCBI Accessession XP_065560174) and in other arthropods (~3057 aa in Anopheles albimanus NCBI Accessession XP_035787416). We repeated the analysis including only 1 of the 4, resulting in a total of 189 transcripts. As the pathways enriched in the specific cell types can bias the results of the GO analysis, we used all the genes expressed for the Germ cells A cluster (3160 genes) as the background for the enrichment analysis. We see an enrichment of terms related to hormone and ion transport across the membrane (Figure 3 E,F, and G). Most of the enrichment terms are associated with the same three genes: nitric-oxide synthase (NOS1), and two different inositol 1,4,5-triphosphate receptor (IP3R) genes (one located on Chromosome_1 and the other on Chromosome_5). While these enrichments do not necessarily correspond to different functions, and it is possible that only one is biologically relevant, it highlighted these three genes as promising candidates.

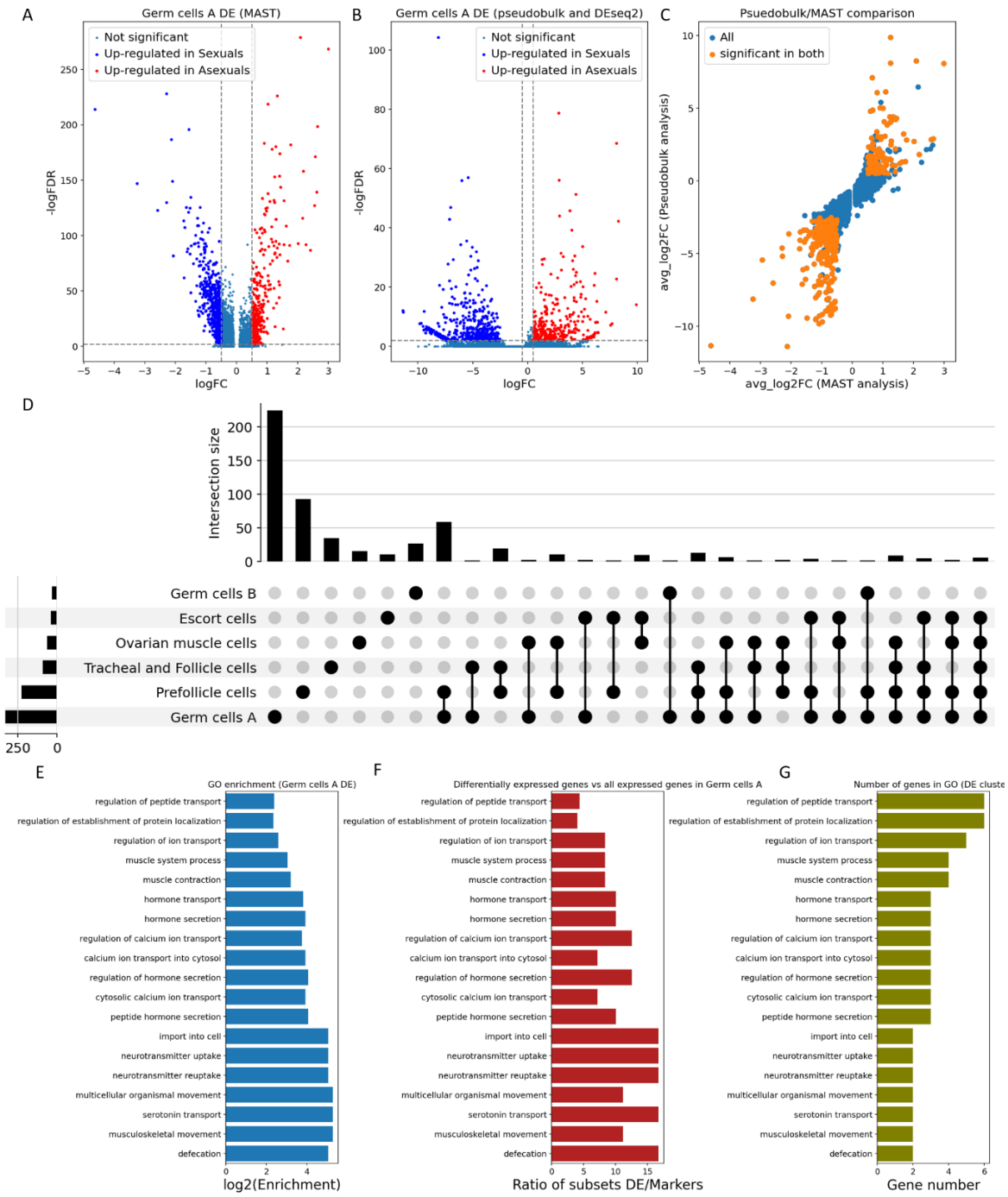


Figure 3: Germ cells A have the largest number of differentially expressed genes. A) Volcano plot showing the results of the DE analysis using MAST in germline cells A. B) Volcano plot showing the results of the DE analysis using the pseudo-bulk approach and DEseq2 in germline cells. C) scatter plot comparing the log2 fold changes from the two analyses (MAST vs psuedo-bulk). D) Upset plot showing the total number of genes significant in both analyses in each cluster, along with the number of intersections between all the sets. E) log2(GO enrichment) in the significant DE genes in germ cells A. F) The ratio of enrichment of the DE genes in germ cell A compared to the enrichment in all the genes expressed in the cluster. G) Number of DE genes in each GO term.

SNPeff analysis and coverage patterns suggest that IP3R is highly degenerated or entirely missing from the *Artemia parthenogenetica* Z

As most of the changes at the transcriptional-level are likely a consequence of a single or multiple genetic variants likely to be present on the Z chromosome, we wanted to identify variants that cause disruptions to protein coding genes in the asexual species. To do so, we mapped the WGS and snRNAseq reads of the two species to the *A. sp. Kazakhstan* genome

and then called SNPs. We filtered for SNPs that were homozygous in *A. sp. Kazakhstan* (0/0) and (0/1 or 1/1) in *Artemia parthenogenetica*. We then used SNPeff to look for variants with high impact in the region with elevated Fst on chromosome 1 (starting from 30,000,000 bp), (section Backcrossing experiment provides further evidence for a Z–linked asexuality locus). We identified 26 variants (Supplementary Table 2). Three of those high impact variants were in the inositol 1,4,5-trisphosphate receptor-like (IP3R) gene that was identified in the last section, and two were predicted to be splice acceptor variants. In order to check if they actually disrupt splicing, we extracted all the reads belonging to germ cells A from the bam files and used HTseq to count the reads mapping to each exon and DEXseq to perform differential exon usage analysis. We identified multiple exons that were upregulated in *A. sp. Kazakhstan* (Figure 4 A).

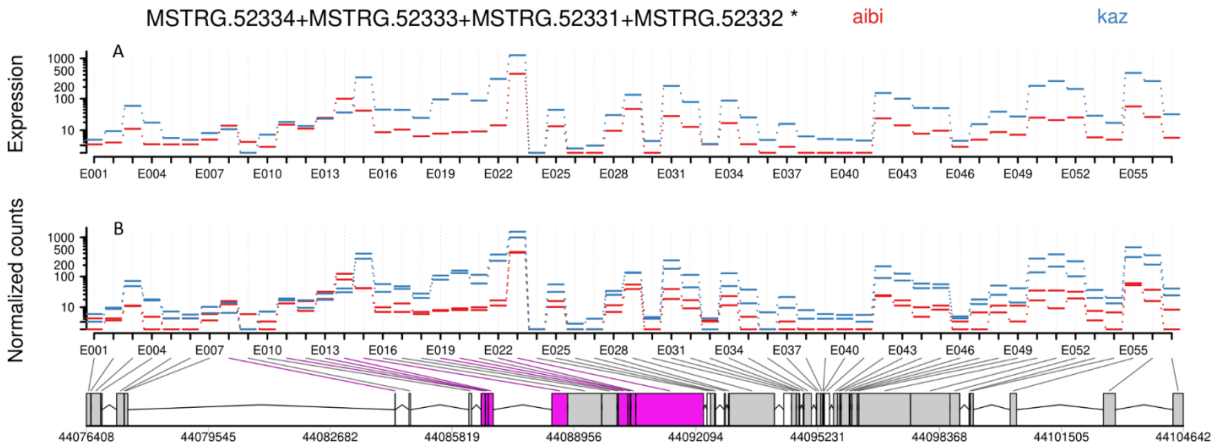


Figure 4: Differential exon usage analysis using the germline A reads highlights extreme differences in alternative splicing patterns. A) Expression values of the Germ cell clusters from the two species across one of the IP3R fragments. B) Normalized counts of the two replicates of each species. The purple colored exons are significant.

While differences in exon usage may reflect point changes at splicing sites, they can also be caused by deletions or duplications of (parts of) the gene in one lineage. To check for copy number variations between the two species in the region containing IP3R, and to identify other potential rearrangements of interest in the genomic region with high Fst between the lineages, we estimated the coverage across the Z chromosome for both species in windows of 10,000 bp, and plotted the ratio of male to male and female to female coverage between the two species and the female/male coverage for both. IP3R falls within the differentiated part of the Z-chromosome, suggesting that it was ancestrally lost from the W. It also has lower coverage in *Artemia parthenogenetica* compared to *Artemia sp. Kazakhstan* in both the male and female comparisons. This reduced coverage suggests that it was (at least) partly lost from the *A. parthenogenetica* Z chromosome, which would explain its pattern of expression and splicing. Another smaller deletion seems to be present upstream of IP3R (at ~43Mb). This region contains another gene that could potentially also be contributing to the shift to asexuality. However, the fact that IP3R is consistently detected across our expression and genome analyses currently makes it the most promising candidate to follow up on.

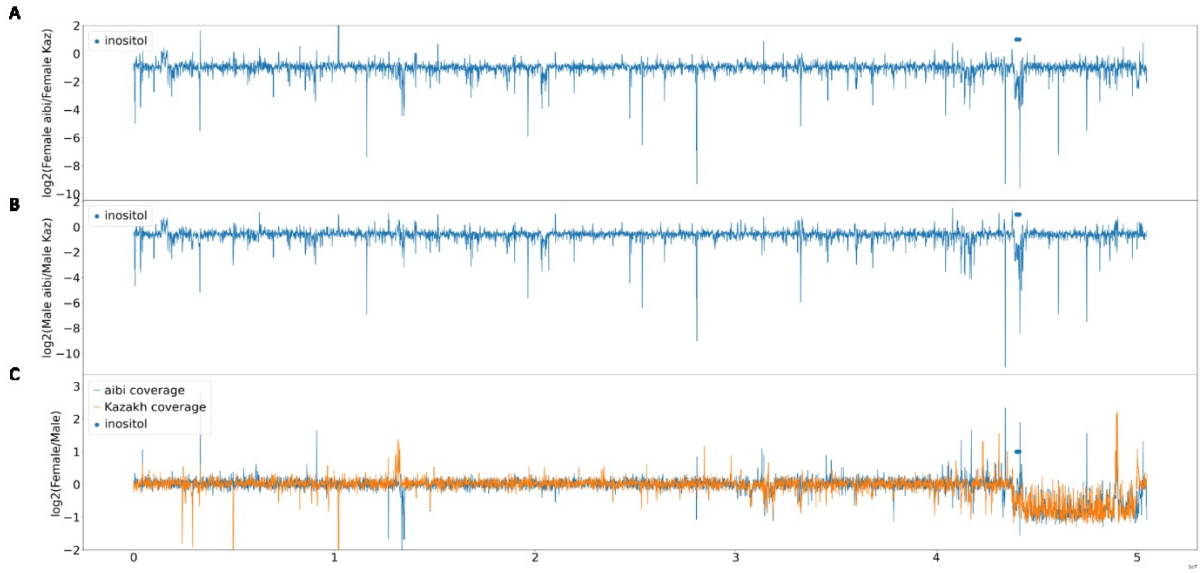


Figure 5: genomic coverage patterns highlight the region containing IP3R is degenerated in *Artemia parthenogenetica* (Aibi lake). A) Log2(Female aibi/Female Kaz) coverage. B) Log2(male aibi/male Kaz) coverage. C) log2(Female/male) coverage for the two species.

Discussion

Z chromosome is the vehicle for contagious parthenogenesis is two asexuals

The distribution of sexual and asexual lineages throughout the tree of life, and in particular the predominance of the former in animals, have been a long term topic of discussion in evolutionary biology. Much of the focus has been on understanding what selective disadvantages might prevent asexual lineages from taking over given their obvious mechanistic advantage. The other side of these dynamics is how often and easily new asexual lineages can be created from sexual ones. The investigation of these transitions has been greatly facilitated by the recent production of extensive genomic and transcriptomic resources in clades containing both. Such studies have allowed both the study of the relationship between sexuals and asexuals (13), and the identification of a few loci underlying the shift (14–16). Evidence from flow cytometry and nuclear DNA data suggested that all the *Artemia parthenogenetica* species shared a common ancestor and likely arose within the last 80,000 years (13). We have previously published data comparing F2 asexual and control females and suggesting the inheritance of a specific region of the Z-chromosome from a rare male (Aibi population) was consistent with the ability to reproduce asexually (10). We repeated the experiment using rare males from a different parthenogenetic population (Atanasvosko) and sequenced more individuals (9 asexual and 9 control). We recover the same pattern, with the only consistent signal across all the chromosomes coming from the Z-chromosome (Figure 1), which does suggest that the Z chromosome likely carries the asexuality locus and that it is the same between the two different populations, further supporting the argument for a common origin for all asexuals.

Early germline cells show the most extreme patterns of differential expression

Previous work by our lab characterizing gene expression in ovaries sexual and asexual *Artemia* lineages found very few systematic differences between the two. While this is in line with automixis using largely the same cellular machinery as sexual meiosis, we reasoned that biologically meaningful differences may only occur in specific cell types, and be difficult to identify in bulk RNA. In particular, cytological evidence suggested that the oocytes of *Artemia* franciscana and asexual *Artemia* show the most differences in chromosomal morphology during middle oogenesis, which is when the oocytes move to the oviducts (7). This is also the stage where differential expression of the meiotic recombination and double-stranded breaks (DSB) genes, Dmc1 and RAD51, was observed (7). Our single nucleus analysis allowed us to focus on cells of the germline, where these changes should be occurring. Although neither of

those genes show significant expression differences between the two species in our analysis (Supplementary Figure 5), both are primarily expressed in germ cells A, the earliest germ cells we detect, which is also where we find the largest number of differentially expressed genes (Figure 3 D). This is in line with the cytology, which suggested that suppression of meiosis I division is the key difference between sexual and asexual species. GO analysis of the differentially expressed genes recovered terms related to the transport of peptides, ions and hormones across the membrane. This is also consistent with the GO analysis comparing the differentially expressed genes between an asexual and A. franciscana, with most of the genes recovered being associated with membrane (1st in cellular component) and transporter activity (3rd in molecular function) (7). This may reflect the fact that cell division requires extensive modulation of membrane components, some of which will be unnecessary in the absence of a meiosis I division.

A possible role for the inositol triphosphate pathway genes in the regulation of asexual meiosis in *Artemia*

The detection of IP3R as our main candidate for the switch to asexuality may at first sight seem surprising, as it is not typically characterized as a meiotic gene. However, the regulation of intracellular calcium, in which it is involved, has emerged as a crucial mechanism for correct oogenesis progression. In particular, during oogenesis, the egg activation is triggered by an increase in calcium levels in all the species studied so far (17). In many species, including mice, this is triggered by fertilization, which activates the inositol triphosphate pathway and results in the release of the calcium stored in the endoplasmic reticulum (18). In *Drosophila* and other insects, on the other hand, the egg activation is triggered by mechanical stimulus without fertilization as the eggs progress through the reproductive tract, but it still requires the increase in calcium (18). The inositol 1,4,5-trisphosphate receptor (IP3R), which codes for an intracellular calcium-release channel, is expressed in the *Drosophila* ovary (18), and although it might not be required for the initiation of the calcium wave, it seems to be required for its propagation (19). IP3R and calcium signalling seem to also be involved in ecdysone secretion in the prothoracic gland in *Drosophila* (20). In *Xenopus*, IP3R seems to be essential for spindle assembly and emitting the first polar body (21). In our results, we observe a deficiency of 2 different IP3R-like genes in the asexual species, with the one located on the Z chromosome overlapping a degenerated region (Figure 5). The upregulation of an IP3R gene in the gonads of the sexual species of *Artemia* was also detected as in the previous analysis using bulk RNAseq (8). As parthenogenetic *Artemia* do not arrest during meiosis and do not divide after meiosis I (7), the lack of a functional copy of IP3R/deficiency and a disruption or modification of the calcium signalling pathway could be one of the changes that facilitated the shift to the asexual mode of reproduction. In mice, fertilization also initiates the degradation of IP3R (17). As it is not clear when the Drosophila-like regulation of egg activation evolved and what role IP3R plays in other arthropod species, more studies/experiments would be necessary to understand its essentiality for crustacean meiosis and what other adaptations were necessary to evade the negative consequences of its deficiency in parthenogenetic *Artemia*.

Methods

Backcrossing experiment and whole-genome sequencing

We mated three rare males from the *Artemia parthenogenetica* (Atanasovsko population) with females from *Artemia sp. Kazakhstan*. The F1s were split into vials not long after their birth, and 8 F1 males were collected to backcross to A. sp. Kazakh females. The F2 were isolated into individual vials, and 9 females which reproduced asexually were frozen instantly upon finding nauplii. After three months, 9 of the remaining females which did not produce any offspring were frozen as control females. Genomic DNA was extracted from the 18 samples using the Qiagen DNeasy Blood & Tissue kit and sent to the Vienna BioCenter (VBC) for sequencing.

Genome Assemblies and Annotation

We assembled the genomes of the two species from the published short read genomic data (8,10,22) using megahit and scaffolded the genomes with the same reads using SOAPdenovo-fusion. The generated scaffolds were anchored with RagTag (23) using the *Artemia sinica* chromosome-level genome assembly as the reference. Each species' genome was annotated by mapping the published short read RNAseq data to the genome using HISAT2 (24) and then generating GTFs with StringTie2 (25).

SNP calling, Local PCA analysis, and Fst analyses

We aligned the WGS reads from the two backcrossing experiments to the A. sp Kazakhstan genome using BWA-MEM (26). We called the SNPs using bcftools for all the samples along with WGS samples for one male and one female from A. sp. Kazakhstan and *A. parthenogenetica*, and pooled RNAseq samples (aligned using HISAT2 (24)) from *Artemia parthenogenetica* (Atanasovsko) females and rare males. The local PCA analysis was run using the lostruct package (27). Fst was calculated using vcftools (--weir-fst-pop) (28).

Nuclei isolation and single-nucleus RNA sequencing

We washed adult *Artemia* females in MilliQ water for 15 minutes, and then we dissected the female reproductive organs under a stereomicroscope and collected them in PBS and placed them on ice. After dissecting, we followed the same protocol described in (12,29) to isolate nuclei and then sent the samples to the Vienna BioCenter sequencing facility for sorting, library preparation and 10x 3' GEX single cell sequencing. We generated four samples with 25 females each: two from *Artemia parthenogenetica* (Aibi Lake) and two from *Artemia sp. Kazakhstan*.

snRNAseq data analysis

The single nucleus RNAseq reads were aligned to the genomes using CellRanger (30) with the GTFs described in (section Genome Assemblies and Annotation). The resulting count matrices were corrected using CellBender (31). The corrected matrices were loaded into Seurat (32) and filtered for nCount_RNA >250. The detection of the highly variable genes was performed using DUBStepR (33), and the integration of the different replicates was performed using Harmony (34).

Differential exon usage analysis

We used the subset-bam tool (https://github.com/10XGenomics/subset-bam) to extract the alignments belonging to the Germ cells A cluster from the different bam files. We then used the dexseq_count.py script, which uses HTseq (35), to get counts for each exon based on the *Artemia sp. Kazakhstan* GTF. The differential exons usage analysis and the plotting were then performed using DEXSeq (36).

References

1. Wilkins AS, Holliday R. The evolution of meiosis from mitosis. Genetics. 2009 Jan;181(1):3–12.
2. Marston AL, Amon A. Meiosis: cell-cycle controls shuffle and deal. Nat Rev Mol Cell Biol. 2004 Dec;5(12):983–97.
3. Lenormand T, Engelstädter J, Johnston SE, Wijnker E, Haag CR. Evolutionary mysteries in meiosis. Philos Trans R Soc B Biol Sci [Internet]. 2016 Oct 19 [cited 2024 Feb 14]; Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0001
4. Bell G. The Masterpiece of Nature :: The Evolution and Genetics of Sexuality. CUP Archive; 1982. 654 p.
5. Engelstädter J. Asexual but Not Clonal: Evolutionary Processes in Automictic Populations. Genetics. 2017 Jun;206(2):993–1009.
6. Nougué O, Rode NO, Jabbour-zahab R, Ségard A, Chevin LM, Haag CR, et al. Automixis in *Artemia*: solving a century-old controversy. J Evol Biol. 2015 Dec;28(12):2337–48.
7. Xu LY, Wu WT, Bi N, Yan ZJ, Yang F, Yang WJ, et al. A cytological revisit on parthenogenetic *Artemia* and the deficiency of a meiosis-specific recombinase DMC1 in the possible transition from bisexuality to parthenogenesis. Chromosoma. 2023 Jun 1;132(2):89–103.
8. Huylmans AK, Macon A, Hontoria F, Vicoso B. Transitions to asexuality and evolution of gene expression in *Artemia* brine shrimp. Proc Biol Sci. 2021 Sep 29;288(1959):20211720.
9. Maccari M, Gómez A, Hontoria F, Amat F. Functional rare males in diploid parthenogenetic *Artemia*. J Evol Biol. 2013 Sep 1;26(9):1934–48.
10. Elkrewi M, Khauratovich U, Toups MA, Bett VK, Mrnjavac A, Macon A, et al. ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp. Dyer K, editor. Genetics. 2022 Sep 30;222(2):iyac123.
11. Boyer L, Jabbour-Zahab R, Joncour P, Glémin S, Haag CR, Lenormand T. Asexual male production by ZW recombination in *Artemia parthenogenetica*. Evolution. 2023 Jan 23;77(1):1–12.
12. Elkrewi M, Vicoso B. Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome. PLOS Genet. 2024 Aug 30;20(8):e1011376.
13. Rode NO, Jabbour-Zahab R, Boyer L, Flaven É, Hontoria F, Van Stappen G, et al. The Origin of Asexual Brine Shrimps. Am Nat. 2022 Aug;200(2):E52–76.
14. Eads BD, Tsuchiya D, Andrews J, Lynch M, Zolan ME. The spread of a transposon insertion in Rec8 is associated with obligate asexuality in *Daphnia*. Proc Natl Acad Sci. 2012 Jan 17;109(3):858–63.
15. Jaquiéry J, Stoeckel S, Larose C, Nouhaud P, Rispe C, Mieuzet L, et al. Genetic Control of Contagious Asexuality in the Pea Aphid. PLOS Genet. 2014 Dec 4;10(12):e1004838.
16. Yagound B, Dogantzis KA, Zayed A, Lim J, Broekhuyse P, Remnant EJ, et al. A Single Gene Causes Thelytokous Parthenogenesis, the Defining Feature of the Cape Honeybee *Apis mellifera capensis*. Curr Biol. 2020 Jun 22;30(12):2248-2259.e6.
17. LEE B, YOON SY, MALCUIT C, PARYS JB, FISSORE RA. Inositol 1,4,5-Trisphosphate Receptor 1 Degradation in Mouse Eggs and Impact on [Ca2+]i Oscillations. J Cell Physiol. 2010 Jan;222(1):238–47.
18. Sartain CV, Wolfner MF. Calcium and egg activation in *Drosophila*. Cell Calcium. 2013 Jan;53(1):10–5.
19. Kaneuchi T, Sartain CV, Takeo S, Horner VL, Buehner NA, Aigaki T, et al. Calcium waves occur as *Drosophila* oocytes activate. Proc Natl Acad Sci. 2015 Jan 20;112(3):791–6.
20. Yamanaka N, Marqués G, O'Connor MB. Vesicle-Mediated Steroid Hormone Secretion in *Drosophila* melanogaster. Cell. 2015 Nov 5;163(4):907–19.
21. Li R, Ren Y, Mo G, Swider Z, Mikoshiba K, Bement WM, et al. Inositol 1, 4, 5-trisphosphate receptor is required for spindle assembly in *Xenopus* oocytes. Mol Biol Cell.

2022 Dec 1;33(14):br27.
22. Huylmans AK, Toups MA, Macon A, Gammerdinger WJ, Vicoso B. Sex-Biased Gene Expression and Dosage Compensation on the *Artemia franciscana* Z-Chromosome. Mank J, editor. Genome Biol Evol. 2019 Apr 1;11(4):1033–44.
23. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022 Dec 15;23(1):258.
24. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug;37(8):907–15.
25. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019 Dec 16;20(1):278.
26. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013 [cited 2023 Sep 4]; Available from: https://arxiv.org/abs/1303.3997
27. Li H, Ralph P. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. Genetics. 2019 Jan 1;211(1):289–304.
28. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.
29. McLaughlin CN, Qi Y, Quake SR, Luo L, Li H. Isolation and RNA sequencing of single nuclei from *Drosophila* tissues. STAR Protoc. 2022 Jun 17;3(2):101417.
30. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8(1):14049.
31. Fleming SJ, Chaffin MD, Arduini A, Akkad AD, Banks E, Marioni JC, et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat Methods. 2023 Sep;20(9):1323–35.
32. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2023 May 25;1–12.
33. Ranjan B, Sun W, Park J, Mishra K, Schmidt F, Xie R, et al. DUBStepR is a scalable correlation-based feature selection method for accurately clustering single-cell data. Nat Commun. 2021 Oct 6;12(1):1–12.
34. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec;16(12):1289–96.
35. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–9.
36. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012 Oct;22(10):2008–17.

Supplementary material

Supplementary tables

Supplementary Table 1: Number of cells assigned to the different clusters in the different replicates (*Artemia sp. Kazakhstan* analysis)

Seurat cluster	Annotation	Kaz replicate 1	Kaz replicate 2	Aibi replicate 1	Aibi replicate 2
0	Escort cells	304	1542	2092	2981
7		5	261	147	228
1	Tracheal and Follicle cells	449	1922	728	1699
2	Germ cells A	627	1885	445	737
3	Ovarian muscle cells	248	1376	333	694
4	Germ cells B	98	1437	67	284
5	Prefollicle cells	156	803	202	315
6		91	667	87	281

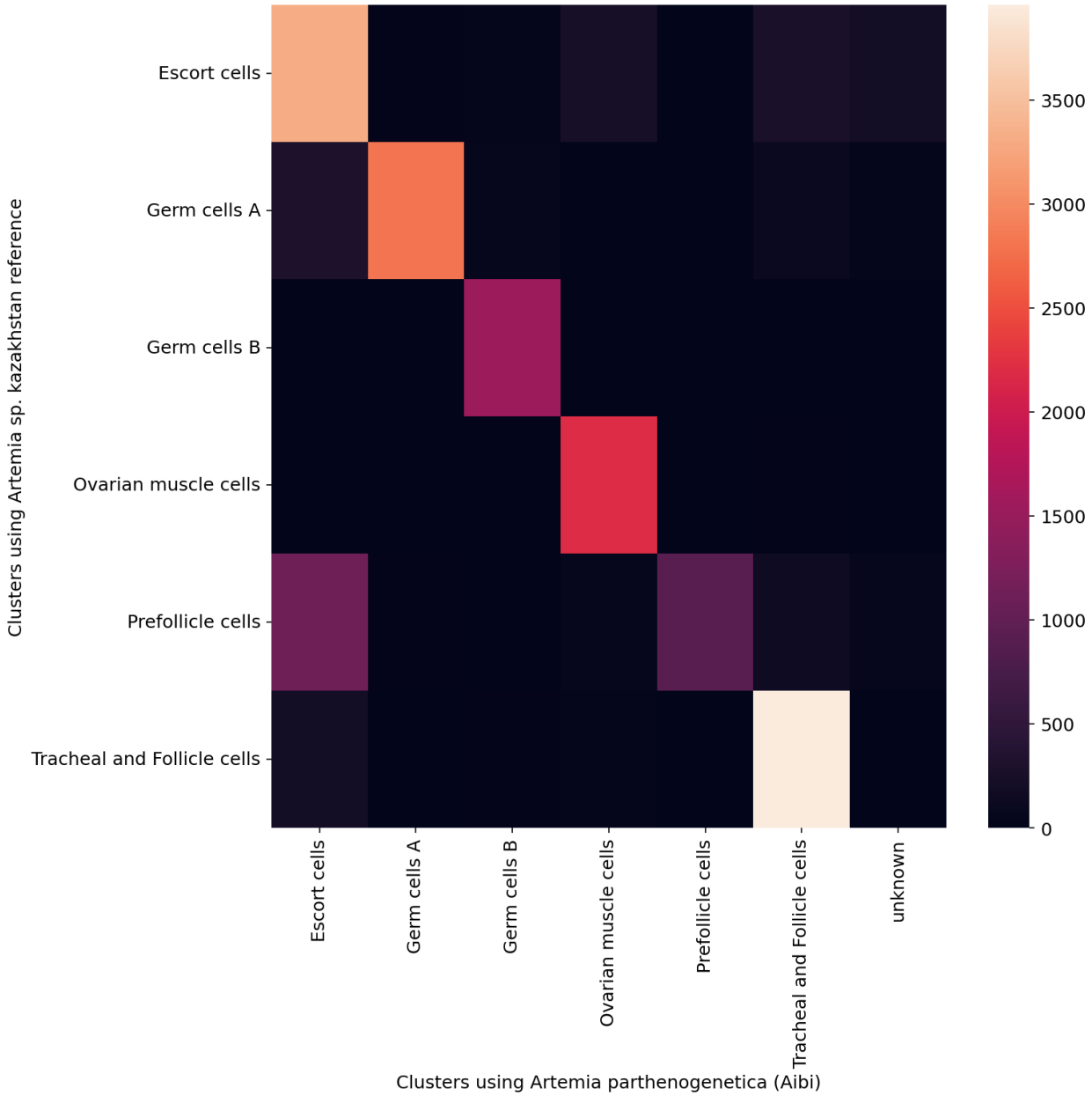
Supplementary Table 2: Number of cells assigned to the different clusters in the different replicates (*Artemia parthenogenetica* analysis)

	Kaz replicate 1	Kaz replicate 2	Aibi replicate 1	Aibi replicate 2
0	559	1317	2234	1027
1	502	2005	722	1351
2	630	1478	312	478
3	298	1408	384	490
4	144	1323	99	243
5	90	495	84	250
6	163	203	16	8

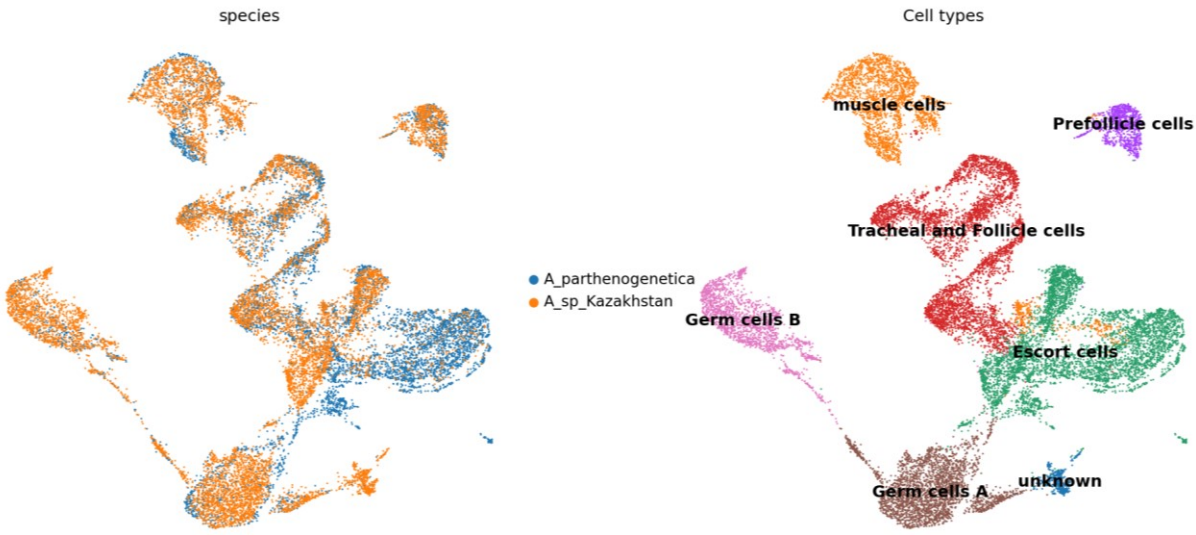
Supplementary Table 3: List of the high impact variants on Chromosome_1 (> 30,000,000bp) and their effects.

Chr	Position	Ref	Alt	genotype	effect	genes
1	30674494	G	A	0/1	stop_gained	MSTRG.51645, MSTRG.51657
1	31085230	G	T	0/1	stop_gained	MSTRG.51675
1	34096679	G	A	0/1	stop_gained	MSTRG.51827
1	34233250	T	C	1/1	splice_donor_variant+intron_variant	MSTRG.51830
1	34237530	G	A	1/1	splice_donor_variant+intron_variant	MSTRG.51830
1	35111456	C	T	0/1	stop_gained	MSTRG.51876
1	35136761	C	A	0/1	splice_donor_variant+intron_variant	MSTRG.51877, MSTRG.51876
1	35383660	G	C	1/1	splice_acceptor_variant+intron_variant	MSTRG.51881, MSTRG.51876
1	35942254	TGG	TG	1/1	frameshift_variant	MSTRG.51895, MSTRG.51896
1	37549323	C	A	0/1	start_lost	MSTRG.51986
1	41578053	A	T	0/1	stop_gained	MSTRG.52183
1	41578063	T	G	0/1	start_lost	MSTRG.52183
1	41855372	G	T	0/1	stop_gained	MSTRG.52202
1	42468505	C	A	0/1	stop_gained	MSTRG.52233
1	42593747	C	A	0/1	stop_gained	MSTRG.52247
1	43046998	T	TA	0/1	frameshift_variant+splice_region_variant	MSTRG.52261

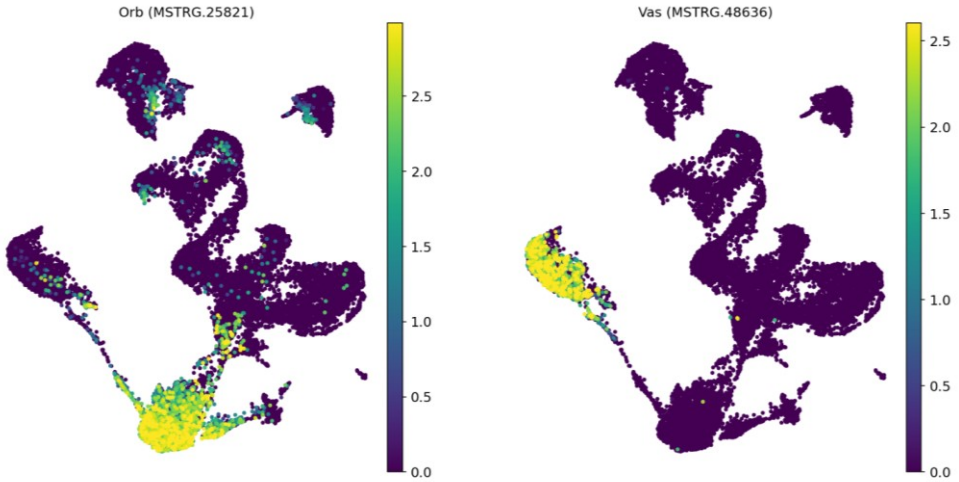
1	**43108737**	**C**	**T**	**0/1**	**splice_donor_variant+intron_variant**	**MSTRG.52263**
1	**43419699**	**C**	**T**	**0/1**	**stop_gained**	**MSTRG.52271**
1	**44090419**	**A**	**T**	**1/1**	**splice_acceptor_variant+intron_variant**	**MSTRG.52331**
1	**44090420**	**C**	**T**	**1/1**	**splice_acceptor_variant+intron_variant**	**MSTRG.52331**
1	**44156071**	**C**	**T**	**0/1**	**stop_gained**	**MSTRG.52337**
1	**44415784**	**G**	**A**	**1/1**	**stop_gained**	**MSTRG.52364, MSTRG.52362**
1	**46039442**	**A**	**T**	**1/1**	**stop_gained**	**MSTRG.52441**
1	**47361817**	**AAAG AAG**	**AAAG**	**0/1**	**splice_acceptor_variant+splice_region_variant+intron_variant**	**MSTRG.52497**
1	**50151222**	**G**	**A**	**1/1**	**stop_gained**	**MSTRG.52637**
1	**50158513**	**G**	**T**	**1/1**	**splice_donor_variant+intron_variant**	**MSTRG.52637**



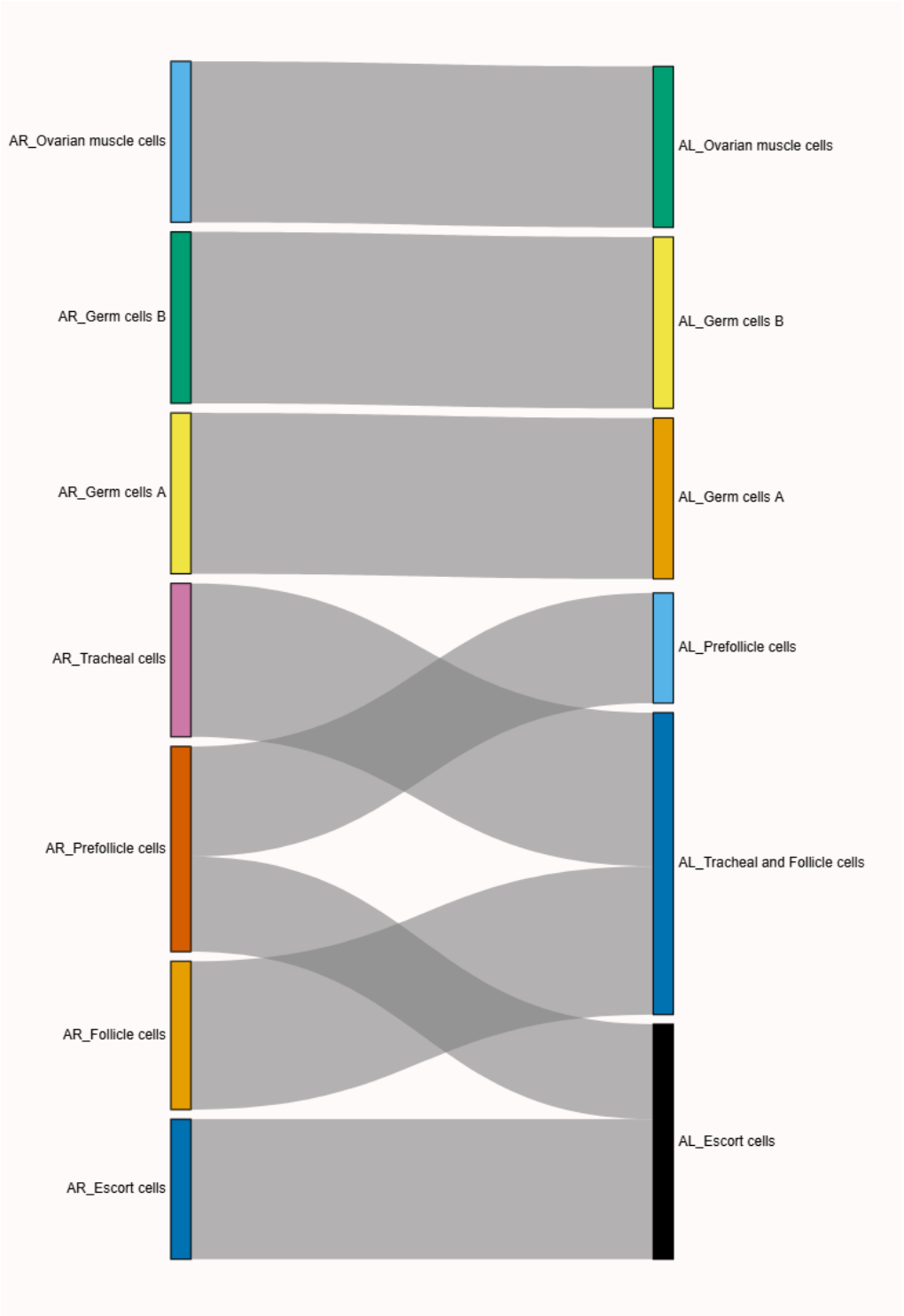
Supplementary Figure 1: Heatmap of the barcode assignments to the different cells type when using *Artemia sp. Kazakhstan* and when using *Artemia parthenogenetica* as references.



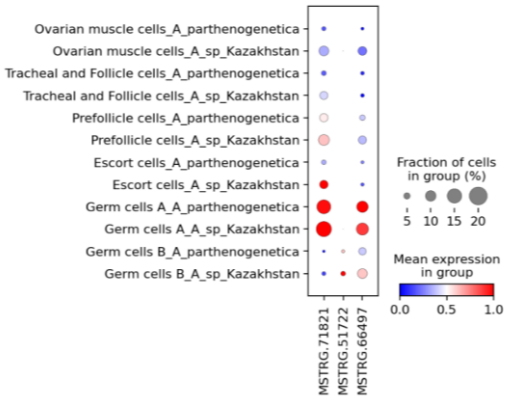
Supplementary Figure 2: A) UMAP colored by species. B) UMAP colored by the cell type annotations transferred from the *Artemia franciscana* atlas (Using *Artemia sp. Parthenogenetica* as reference).



Supplementary Figure 3: A) The expression of the early germlice markers (orb). B) The expression of the late germline marker (vas)



Supplementary Figure 4: A Sankey diagram showing the correspondence between the clusters from this analysis (Aibi reference) and the *Artemia franciscana* atlas.



Supplementary Figure 5: The expression of Dmc1 and RAD51 (2 hits) orthologs in the different cell types.

Chapter 6: Thesis discussion

Artemia, a new model for studying sex chromosome evolution

In recent years, large-scale genomics paved the way for the rapid identification of sex chromosomes in a large variety of invertebrate and vertebrate animals and plants. This has yielded new insights into sex chromosome evolution, such as evidence for a larger number of sex chromosome turnovers than previously anticipated (1–5). Despite this abundance of data, we still lack convincing evidence to support mechanistic and evolutionary explanations that have been proposed for several of the early steps of sex chromosome evolution. In particular, the forces that promote the progressive loss of recombination, and the molecular mechanisms underlying these changes, have still to be systematically understood and remain under vigorous debate (6–9).

Our results in *Artemia* Brine Shrimp make this clade a promising model to tackle these questions. While the ZW pair is ancestral to the clade, new strata have formed independently and convergently around it in the American and Eurasian lineages (Figure 1). One known mechanism to achieve reduced recombination is the fixation of inversions on either the Z or W, as these prevent pairing and crossing over in heterozygous individuals. While a complete assembly of the W chromosomes of the two lineages is still missing, future improved assemblies will make it possible to search for inversions between the Z and W chromosomes in the different lineages that might have contributed to the loss of recombination. Changes in the epigenetic landscape by methylation and histone modifications have also been suggested as a possible culprit in the suppression of recombination (10). One possible way to test that hypothesis in *Artemia* will be to compare the epigenetic landscapes of the Z and W chromosomes in the strata of different ages across species. This might give some insights into the presence/absence of progressive changes in the methylation patterns and the enrichments of different histone modifications.

Additionally, a deeper look at the gene content in the non-recombining regions, in terms of what genes get captured and the events of gene duplications, additions and deletions, might give insights about the forces that play a role in promoting the loss of recombination. Selective hypotheses typically invoke the advantage of keeping female-beneficial alleles on the female-specific W chromosome (11). In theory, this should lead to the evolution of at least one gene that becomes specialized for female-specific functions on each stratum (while the ancestral Z copy may maintain its original function, become specialized for male function, or be lost entirely), likely under the action of positive selection. On the other hand, genes with such specialized female-specific functions may not appear in the early stages of W evolution if loss of recombination is neutral (12), or represents the fixation of a “lucky inversion” that happened to carry fewer deleterious mutations than other haplotypes in the population, thereby obtaining a selective advantage (13).

Another open question involves the direction of causality between degeneration of W-linked genes and dosage compensation of their Z homologs. Although the traditional model posits that dosage compensation evolves in response to degeneration (14), a new model recently proposed that a runoff regulatory process that concurrently leads to decreased expression of W-linked genes and increased expression of Z-linked genes (i.e. dosage compensation) drives (and therefore predates) the recombination arrest (15). It might be possible to test the predictions of this theory by checking if pseudogenized genes in the younger strata already show evidence of dosage compensation, and divergence in their cis-regulatory sequences. Such analyses in various ZW and XY systems of various ages will allow for a more quantitative testing than has so far been possible of the various hypotheses that can (individually or together) shape the fate of new sex determining regions.

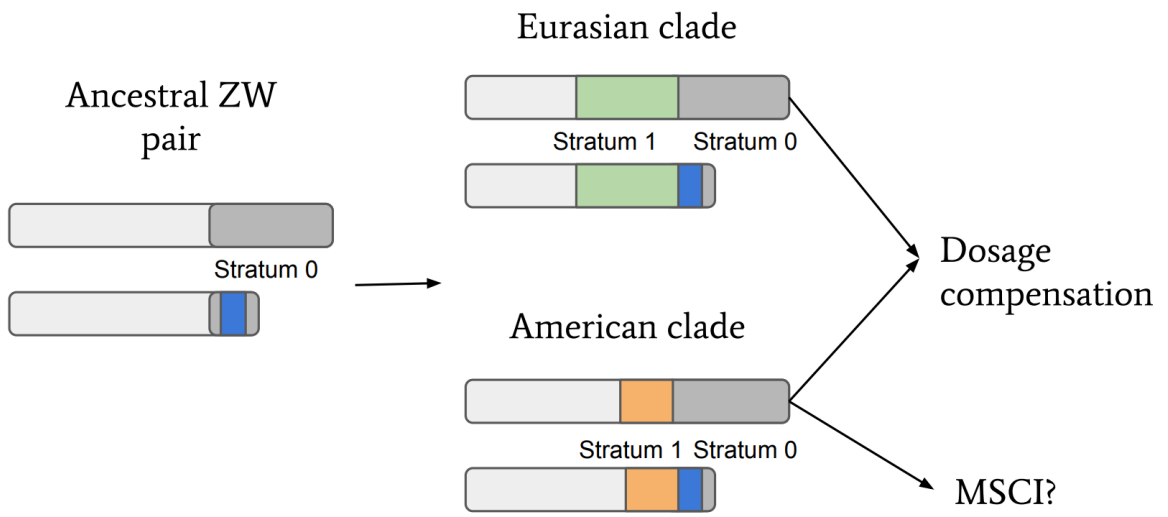


Figure 1: The ancestral and young strata on the Z and W chromosomes in the American and Eurasian lineages.

Some sex determining genes are easier to find than others

The extensive sex chromosome turnover observed in some clades implies that novel sex determining loci must arise frequently, and understanding how novel sex determining genes arise has been an important goal. While much progress has been made since the advent of next generation sequencing and gene editing technologies, our understanding of this process suffers from a historical bias, because the strategies used for the identification of different types of sex determination genes have varying degrees of difficulty. The simplest approach is to look for paralogs of previously identified sex determination genes in Y- or W-linked regions. This method was successfully used to find most of the vertebrate sex determination genes outside of the model species (16). The fact that many such paralogs have been found has led to the "usual suspects" hypothesis, the idea that constraints in the evolution of the sex determining pathway only allow for genes that are already in the cascade to take over the primary determinant role.

The second approach that yields relatively straightforward results is to look at the few functional genes retained on the ancient part of the Y/W chromosomes. This strategy was used successfully to identify the Y-linked sex determination gene in the Mediterranean fruit fly, MOY (17), and the Y-linked gene, YOB, controlling the splicing of Dsx and fru in Anopheles Mosquitoes (18). Neither of these is related to known sex determining genes of other clades. We have also used this approach to identify a W-linked splicing factor (u2af2) as a promising candidate sex-determination gene in schistosomes (19). Systems where non-coding genes are involved required more complex strategies to identify and validate candidate sex-determination genes. For instance, in Bombyx mori, sex turned out to be determined by a W-linked piRNA precursor, the identification of which involved using more elaborate approaches, including sexed RNAseq data of embryos, piRNA data, RNAi and piRNA inhibitors (20). Finally, sex determination mechanisms that involve one or multiple dosage dependent genes are perhaps the most complex to tackle, such as the *Drosophila* and *C. elegans* cases, where mutational screens were necessary to identify the key players in the cascade (21,22).

While this diversity of systems has clearly demonstrated that there is more to the evolution of sex determination than the usual suspects hypothesis, understanding how often such deviations occur (and why) will require the systematic detection of primary sex determining signals that are difficult to find. In *Artemia*, which belongs to a clade where sex is largely determined by environmental factors (23), we have attempted to identify the master gene using a combination of genomic and transcriptomic approaches. Although we have managed to highlight some of candidate genes in the sex determination pathway (Chapter 3), the upstream signal remains elusive. This could be due to the regulator being a dosage dependent gene on the Z expressed only transiently during early development, or to it belonging to other

classes of functional molecules in the cell, including long non-coding RNAs (without a poly-A tail) and/or small RNAs, for which we have no data so far. To tackle those challenges, our lab is performing an expansive search for coding and non-coding sex-linked transcripts on the W, including the generation of small RNA libraries for males and females and better assemblies of the W chromosome and its gene content. The hope is that the findings and methods developed will provide insights that will translate into better strategies for other complex cases of sex determination.

MSCI: how do we settle a case like *Drosophila* (or *Artemia*)?

In our study of the Z chromosome regulation in germline cells of *Artemia*, we have come across similar challenges to the ones that are driving the debate on the status of MSCI in *Drosophila melanogaster* (24,25). Most of those challenges have biological roots, which are made more complex by the limitations of our current technologies. In the case of *Artemia* and *Drosophila*, the transcriptional readout for the lack of dosage compensation and meiotic sex chromosome inactivation is of the same nature, which is a decrease in the abundance of transcripts originating from the Z chromosome, but potentially different in the extent/degree. As the difference is mainly in the degree, disentangling the two mechanisms using RNAseq technologies becomes very difficult. Bulk RNAseq of whole tissues masks the tissue heterogeneity, and snRNAseq is inherently noisy and provides a very sparse representation of the transcriptomes of each cell. More generally, RNA-seq transcript abundances do not necessarily reflect the current state of transcription in the cell due to the RNA degradation kinetics and other factors, like maternal mRNA storage in oocytes (26,27).
In our analysis, we followed a very stringent approach to the problem and managed to complement the single-cell RNAseq with ATACseq data from the same cells. Observing a similar pattern in the ATACseq data provided some additional support for the idea that an active mechanism to repress the Z may be present. In the future, adding data from different modalities, such as single-cell nascent RNA sequencing (scGRO-seq) (28), which captures actively transcribed genes, and single-nucleus translatome data, which captures the actively translated genes, would provide a more precise estimate of the transcriptional state of each cell/chromosome. A combination of the previous strategies and approaches that capture the chromatin changes that take place in germline cells (29) would shed new light on the regulatory mechanisms at play in *Drosophila* and *Artemia*, where dosage compensation takes place in the heterogametic sex.
Beyond these technical considerations, much of the debate around *Drosophila* and MSCI at large seems to be of a more philosophical nature: what exactly qualifies as inactivation? How much of a reduction in sex linked expression is biologically meaningful, and the result of a specific selective pressure? Do repressive chromatin marks need to be present, or is it sufficient to detect a deficit of active marks? Species where dosage compensation occurs in the homogametic sex (but MSCI is still necessarily expected in the heterogametic sex), such as nematodes and butterflies, may provide an interesting counterpoint for defining baseline definitions that can then be applied more broadly. Such comparative approaches will hopefully eventually contribute to our understanding of who has acquired true MSCI and why.

Why aren't we all asexual?

Various studies of parthenogenesis suggest that the modifications that lead to the evolution of asexuality tend to be disruptive changes in meiosis-essential genes or processes. In *Daphnia pulex*, for instance, a transposon insertion upstream of one of the three copies in the genome of the meiotic cohesin Rec8 and a frameshift mutation seem to be shared across the asexual lineages, making the locus a likely candidate for the transition to obligate asexuality (30). In the Cape Honeybee, a Hymenopteran-specific protein, with 3 non-synonymous variants and a putative role in chromosome segregation, co-segregates with asexuality and is downregulated in asexual bees, potentially causing issues in meiosis (31). Similarly,

hybridization and polyploidy, two scenarios often associated with defects in gametogenesis, are also strongly associated with switches to asexuality. In *Artemia*, we observe a likely disruption in a homolog of a calcium-channel encoding gene that is important for oocyte development and activation in many other species; interfering with the gene in other species caused meiosis-related defects (32).

Some results in mutants of sexual species also support the idea that interfering with different aspects of the meiotic machinery can lead to asexual meiosis. When characterizing a *Drosophila* mutant line for yem1, a gene involved in chromatin remodeling and chromosome behaviour in meiosis I, Meyer et al. (2010) noticed that their females were producing a small number of offspring without paternal genomic contribution (33). Based on findings from the naturally parthenogenetic *Drosophila mercatorum*, a more recent study also showed that increasing the copy number of Polo (a kinase with meiotic and mitotic functions) and decreasing the expression of Desat2 (a desaturase) was enough to induce facultative parthenogenesis in *Drosophila melanogaster* (34). Taken together, these studies show that disturbing different meiotic mechanisms can lead to the emergence of asexual meiosis.

As asexuality can rise by/be induced by many modifications to meiosis-related genes, the rarity of transitions observed in nature is quite intriguing. One possibility is that new instances of parthenogenesis occur more often than we think, but are associated with strong deleterious consequences (such as low fertility), and therefore quickly eliminated. On the other hand, in species that show stable transitions to asexuality, multiple prerequisite conditions likely had to be satisfied before the shift, such as some level of redundancy in components of the meiotic pathways and certain levels of recombination rate. In the case of *Artemia*, the presence of a second paralog of IP3R could have mitigated some of the deleterious effects of the disruption to the other copy, similar to the multiple paralogs of Rec8 in *Daphnia*. In *Drosophila*, which like other insects only has one copy of IP3R, unconditional mutants do not survive past larval stage (34). Additionally, low recombination levels in parthenogenetic species with central fusion automixis slows down the loss of heterozygosity that takes place under this form of asexual reproduction, and low recombination levels in the ancestral sexual species may be a prerequisite for the survival of its derived asexual lineages, as has been proposed in *Artemia* (35). These required pre-existing conditions may limit the number of lineages that transition successfully to asexuality even in the short term. Finally, the presence of short-lived asexual lineages throughout the tree of life suggests that even when the transition is beneficial in the short term, they tend to be short-lived (36). Studies in *Artemia* and other young and old asexual lineages will provide better insights on whether males should be worried or not, and what the prerequisites are for their disappearance.

References

1. Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, et al. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nat Commun. 2018 Oct 5;9(1):4088.
2. Saunders PA. Sex Chromosome Turnovers in Evolution. In: eLS [Internet]. John Wiley & Sons, Ltd; 2019 [cited 2025 Jan 17]. p. 1–8. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0028747
3. Vicoso B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol. 2019 Dec;3(12):1632–41.
4. Wang D, Li Y, Li M, Yang W, Ma X, Zhang L, et al. Repeated turnovers keep sex chromosomes young in willows. Genome Biol. 2022 Sep 23;23(1):200.
5. Wang J, Tao W, Kocher TD, Wang D. Sex chromosome turnover and biodiversity in fishes. J Genet Genomics. 2024 Dec 1;51(12):1351–60.
6. Ponnikas S, Sigeman H, Abbott JK, Hansson B. Why Do Sex Chromosomes Stop Recombining? Trends Genet. 2018 Jul 1;34(7):492–503.
7. Olito C, Abbott JK. The evolution of suppressed recombination between sex chromosomes and the lengths of evolutionary strata. Evolution. 2023 Apr 1;77(4):1077–90.
8. Charlesworth B, Olito C. Making sense of recent models of the “sheltering” hypothesis for recombination arrest between sex chromosomes. Evol Int J Org Evol. 2024 Dec 2;78(12):1891–9.
9. Jay P, Jeffries D, Hartmann FE, Véber A, Giraud T. Why do sex chromosomes progressively lose recombination? Trends Genet. 2024 Jul 1;40(7):564–79.
10. Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, et al. Sex Chromosome Evolution: So Many Exceptions to the Rules. Genome Biol Evol. 2020 Jun 1;12(6):750–63.
11. Rice WR. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. Evolution. 1987;41(4):911–4.
12. Jeffries DL, Gerchen JF, Scharmann M, Pannell JR. A neutral model for the loss of recombination on sex chromosomes. Philos Trans R Soc B Biol Sci. 2021 Jul 12;376(1832):20200096.
13. Jay P, Tezenas E, Véber A, Giraud T. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLOS Biol. 2022 Jul 19;20(7):e3001698.
14. Charlesworth B. Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci. 1978 Nov;75(11):5618–22.
15. Lenormand T, Roze D. Y recombination arrest and degeneration in the absence of sexual dimorphism. Science. 2022 Feb 11;375(6581):663–6.
16. Bertho S, Herpin A, Schartl M, Guiguen Y. Lessons from an unusual vertebrate sex-determining gene. Philos Trans R Soc B Biol Sci. 2021 Jul 12;376(1832):20200092.
17. Meccariello A, Salvemini M, Primo P, Hall B, Koskinioti P, Dalíková M, et al. Maleness-on-the-Y (MoY) orchestrates male sex determination in major agricultural fruit fly pests. Science. 2019 Sep 27;365(6460):1457–60.
18. Scott M. Sex Determination and Dosage Compensation: *femaleless* Is the Link in *Anopheles* Mosquitoes. Curr Biol. 2021 Mar 8;31(5):R260–3.
19. 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. Wilson M, editor. Mol Biol Evol. 2021 Dec 9;38(12):5345–58.
20. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature. 2014 May;509(7502):633–6.
21. Salz HK, Erickson JW. Sex determination in *Drosophila*. Fly (Austin). 2010 Jan;4(1):60–70.
22. Strome S, Kelly WG, Ercan S, Lieb JD. Regulation of the X Chromosomes in *Caenorhabditis elegans*. Cold Spring Harb Perspect Biol. 2014 Mar 1;6(3):a018366–a018366.

23. Ye Z, Bishop T, Wang Y, Shahriari R, Lynch M. Evolution of sex determination in crustaceans. Mar Life Sci Technol. 2023 Feb 1;5(1):1–11.
24. Elkrewi M, Vicoso B. Single-nucleus atlas of the *Artemia* female reproductive system suggests germline repression of the Z chromosome. PLOS Genet. 2024 Aug 30;20(8):e1011376.
25. Wei KHC, Chatla K, Bachtrog D. Single-cell RNA-seq of *Drosophila* miranda testis reveals the evolution and trajectory of germline sex chromosome regulation. PLOS Biol. 2024 Apr 30;22(4):e3002605.
26. Wang C, Liu H. Factors influencing degradation kinetics of mRNAs and half-lives of microRNAs, circRNAs, lncRNAs in blood in vitro using quantitative PCR. Sci Rep. 2022 May 4;12(1):7259.
27. Cheng S, Altmeppen G, So C, Welp LM, Penir S, Ruhwedel T, et al. Mammalian oocytes store mRNAs in a mitochondria-associated membraneless compartment. Science. 2022 Oct 21;378(6617):eabq4835.
28. Tang L. Capturing nascent transcripts in single cells. Nat Methods. 2024 Jul;21(7):1144–1144.
29. Anderson J, Henikoff S, Ahmad K. Chromosome-specific maturation of the epigenome in the *Drosophila* male germline. eLife [Internet]. 2023 Nov 17 [cited 2024 Jan 29];12. Available from: https://elifesciences.org/reviewed-preprints/89373
30. Eads BD, Tsuchiya D, Andrews J, Lynch M, Zolan ME. The spread of a transposon insertion in Rec8 is associated with obligate asexuality in *Daphnia*. Proc Natl Acad Sci. 2012 Jan 17;109(3):858–63.
31. Yagound B, Dogantzis KA, Zayed A, Lim J, Broekhuyse P, Remnant EJ, et al. A Single Gene Causes Thelytokous Parthenogenesis, the Defining Feature of the Cape Honeybee *Apis mellifera capensis*. Curr Biol. 2020 Jun 22;30(12):2248-2259.e6.
32. Li R, Ren Y, Mo G, Swider Z, Mikoshiba K, Bement WM, et al. Inositol 1, 4, 5-trisphosphate receptor is required for spindle assembly in *Xenopus* oocytes. Mol Biol Cell. 2022 Dec 1;33(14):br27.
33. Meyer RE, Delaage M, Rosset R, Capri M, Aït-Ahmed O. A single mutation results in diploid gamete formation and parthenogenesis in a *Drosophila* yemanuclein-alpha meiosis I defective mutant. BMC Genet. 2010 Nov 16;11:104.
34. Sperling AL, Fabian DK, Garrison E, Glover DM. A genetic basis for facultative parthenogenesis in *Drosophila*. Curr Biol CB. 2023 Sep 11;33(17):3545-3560.e13.
35. Rode NO, Jabbour-Zahab R, Boyer L, Flaven É, Hontoria F, Van Stappen G, et al. The Origin of Asexual Brine Shrimps. Am Nat. 2022 Aug;200(2):E52–76.
36. Moreira MO, Fonseca C, Rojas D. Parthenogenesis is self-destructive for scaled reptiles. Biol Lett. 17(5):20210006.

Appendix

Publications not included in this thesis:
*co-first authors

Lasne C*, Elkrewi M*, Toups MA, Layana L, Macon A, Vicoso B. The scorpionfly (*Panorpa cognata*) genome highlights conserved and derived features of the peculiar dipteran X chromosome. Molecular Biology and Evolution. 2023 Dec;40(12):msad245.

Kelemen RK, Elkrewi M, Lindholm AK, Vicoso B. Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. Proceedings of the Royal Society B. 2022 Feb 9;289(1968):20211985.

Elkrewi M*, Moldovan MA*, Picard MA, Vicoso B. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. Molecular Biology and Evolution. 2021 Dec 138(12):5345-58.