# Evolutionary Genomics of Selection Across Haploid and Diploid Life Stages in Plants

by

Meng Yuan

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Graduate Department of Ecology and Evolutionary Biology
University of Toronto

# Evolutionary Genomics of Selection Across Haploid and Diploid Life Stages in Plants

Meng Yuan

Doctor of Philosophy

Department of Ecology and Evolutionary Biology
University of Toronto

2025

## Abstract

The alternation of a diploid and haploid life stage is shared among all land plants. Both life stages share a common genome and overlap considerably in their gene expression, while differing in their ploidy, morphology, development and genetics. Understanding how selection on one life stage affects the other stage is important in answering central evolutionary questions such as the maintenance of genetic variation. In my thesis, I used population genomics to investigate the potential for conflict between plant life stages in both bryophytes and angiosperms. In Chapter 2, I used cis-regulatory variation to test the scope of evolutionary conflict between sexes and life stages based on the degree of shared genetic architecture in the dioecious plant *Rumex hastatulus*. My results revealed a stronger shared genetic basis between sexes than between life stages, indicating a higher potential for sexual conflict than life-stage conflict when under antagonistic selection. In Chapter 3, I tested the population genomic signal of balancing selection generated by the intralocus conflict between life stages in the angiosperm *Rumex hastatulus* and the moss *Ceratodon purpureus*. The genome-wide patterns of diversity statistics were more consistent with concordant selection between life stages in both species instead of widespread antagonistic balancing selection. Model-based tests on balancing selection identified hundreds of candidate genes in both species that are promising for future research. In Chapter 4, I attempted to combine experimental crosses with pooled sequencing of pollen grains

and early seeds in *Rumex hastatulus* to characterize genetic variation for pollen meiotic drive and pollen competition, two key processes in male haploids of angiosperms. However, I did not find clear signals of meiotic drive based on patterns of allele frequencies and technical limitations prevented me from searching for signals of pollen competition based on genomic coverage on sex chromosomes; my analyses suggested most seeds were likely unfertilized ovules. Overall, my thesis highlights the evolutionary consequences of the biphasic life cycles in plants and even though no widespread life-stage conflict was found, future work is needed for investigating selection and the genetic basis of conflict across plant life stages.

# Acknowledgments

It has been a privilege to complete my PhD journey surrounded by so many amazing people. Coming as a first-generation college student, I feel lucky to pursue my passion and contribute to evolutionary biology research, it would not be possible without the support from my mentors, friends and family.

First, I would like to thank my PhD supervisors Stephen Wright and John Stinchcombe for their support on my path of becoming an independent researcher. I first met Stephen and John at the EEB Barrett Fest Symposium in Summer 2018, it was my first time being exposed to plant reproductive system evolution and talking to many established researchers in plant evolution. After that, I was very lucky to work with Stephen and Felix Beaudry on an undergraduate thesis project, which deepened my knowledge of evolutionary genetics in the coolest plant system *Rumex*. I thank John for all the fun office visits and countless conference introductions. I am very grateful for my collaborations with Stuart McDaniel and Spencer Barrett. Stuart introduced me to a whole new world of moss evolution, and he is always excited to talk about research and new ideas. Spencer offered his extensive expertise in plant evolution and also helped me become a better academic writer. I want to thank my committee members Aneil Agrawal and Jacqueline Sztepanacz for their useful advice and constructive feedback on all stages of my research, and Helen Rodd and Asher Cutter for their questions during my appraisal exam. Thank you to Judith Mank and the rest of my defense exam committee for inspiring questions and discussions.

I want to thank the senior members from my labs for their mentoring and support in many aspects of my research: Felix Beaudry, Joanna Rifkin, Haoran Xue, Tia Harrison, Julia Kreiner, Tyler Kent, Damian Hernandez, and thank you to the rest of my big lab family for inspiring discussions and helpful suggestions: Bianca Sacchi, Amanda Peake, Julia Boyle, George Sandler, Zoe Humphries, Georgia Henry, Alice Fairnie, Louisa Bartkovich, Katie Maunder, Mark Hibbins, Solomiya Hnatovska, Kuangyi Xu, Linyi Zhang, Anna O'Brien, Emily Glasgow, Martin Henry, Erin McHugh, Wen-Juan Lan, Cassandre Pyne, and Ting Liu. A huge thanks to my first-year office mate Karl Grieshop for advice and stimulating conversations on my research. Thank you to Megan Bontrager and Micah Freedman on useful feedback during lab meetings. I am very grateful to Baharul Choudhury and Jack Hu for making the best greenhouse and molecular biology lab scientist out of me, to Yunchen Gong for valuable help with server use,

# Table of Contents

# List of Tables

# List of Figures

# List of Appendices

# Chapter 1: Introduction

## Overview

The alternation of generations is a fundamental feature of land plant evolution. Several aspects of the similarity and differences between the two plant life stages offer great opportunities to test central questions in evolutionary genetics of plants. In my thesis, I used population genomics, combined with transcriptomics, association mapping, and experimental crosses to test the evolutionary consequences of the alternating life cycles between the diploid and haploid phases in both bryophytes and angiosperms. I focused on the potential for conflict between life stages when they have potentially diverging evolutionary interests over traits while sharing a common genome. I tested the potential for conflict based on the extent of shared genetic architecture using gene regulatory variation for expression, examined the population genomic signal of intralocus conflict between life stages, and investigated standing variation for pollen competition, the process of male gametophytic competition in angiosperms. In this introductory chapter, I will first outline the main motivation for my research – the alternation of diploid and haploid life stages in angiosperms and bryophytes. I will then summarize the key features of selection on the haploid phase and its consequences on plant evolution. Finally, I will discuss the theoretical predictions on conflict between plant life stages and empirical tests for its genomic signal. Throughout my thesis, plant life stages refer to the haploid and diploid phases of the plant life cycle rather than developmental stages such as juvenile or adult stages.

## Alternating life cycle of land plants

All sexually reproducing eukaryotes alternate their life cycles between a haploid (n) and a diploid (2n) phase (Mable and Otto 1998). In animals, the haploid phase is restricted to the gametes (egg and sperm) that fuse to become the zygote and develop into a new individual. The gamete phase of many animals (e.g., with internal fertilization) occurs completely within the diploid organism. However, in land plants (a clade of plants that are mainly terrestrial), the haploid phase is considered as a separate organism called a gametophyte; plant gametophytes can go through growth, development and produce gametes that fuse and become the diploid phases called the sporophyte generation (Lewis and McCourt 2004; Taylor et al. 2005). The haploid gametophytes and diploid sporophytes may have varying levels of complexity and

independence, contributing to diverse types of life cycles among land plants (Mable and Otto 1998; Qiu et al. 2012). For example, bryophytes (mosses, liverworts, and hornworts) are considered the 'early' land plants that usually live in moist environments (McDaniel 2021). They lack vascular tissues to transport water or nutrients, and their gametophytes are the more complex and longer-living life stage, while their sporophytes rely on gametophytes for nutrients. In vascular plants, like gymnosperms or angiosperms, sporophytes are independent and free-living, while their gametophytes are ephemeral and have fewer cells. The haploid and diploid life stages of land plants have distinctive morphology and development (Qiu et al. 2012), occupy different ecological niches, and their adaptation to environments ensures their success in terrestrial ecosystems compared to their aquatic ancestors (Harrison 2017; Donoghue et al. 2021; Harris et al. 2022). The differences between the two life stages while sharing a common genome can have important evolutionary consequences. Studying how selection in one life stage affects the other stage is central to understanding land plant evolution and offers many opportunities to test fundamental evolutionary genetics questions.

Angiosperms have the most extremely reduced gametophytes and elaborated sporophytes (Qiu et al. 2012). The haploid stage of angiosperm includes ovule (female) and pollen (male), pollen grains compete for fertilizing the ovule, a process known as "certation", has been observed as early as the late 1800s (Darwin 1876; Haldane 1932; Mulcahy and Mulcahy 1987). After pollination, pollen grains germinate on the stigma and grow pollen tubes in the style to transfer the sperm to the ovule; fertilization success is affected by many factors including pollen competitive ability traits and the ecology of interaction between pollen and pistil (stigma, style, ovule) (Williams and Mazer 2016). Studies have shown heritable variation in pollen performance traits such as pollen tube growth rate that affect male siring success, faster-growing pollen tubes are more likely to fertilize the ovule (Walsh and Charlesworth 1992; Beaudry et al. 2020). Competition between female- and male-determinaing (X- and Y-bearing) pollen grains is suggested to cause female-biased sex ratios in plant species with heteromorphic sex chromosomes (Lloyd 1974; Field et al. 2012). Despite the important consequences of pollen competition, the genomic distribution, and extent and genetic basis of genetic variation for pollen competition remain understudied.

# Haploid selection in plant gametophytes

Plant gametophytes express a large proportion of their genome during growth and development, potentially making the haploid gametophytes subject to substantial selection (reviewed in Beaudry et al. 2020). The widespread gametophytic gene expression could lead to a large overlap in gene expression between life stages, providing an opportunity to test how one life stage affects the other. In *Arabidopsis thaliana*, at least 60% of genes are expressed in pollen (Honys and Twell 2003; Walbot and Evans 2003; Borg et al. 2009; Rutley and Twell 2015). In maize, the haploid ovule and embryo sac also share considerable overlap in expression with the diploid seedling (Chettoor et al. 2014). Compared to the angiosperm *Arabidopsis thaliana*, the overlap in expression is even larger in the moss *Funaria hygrometrica*, potentially reflecting the reduction of the gametophytic stage in angiosperms (Szövényi et al. 2011). For example, 85% of genes are expressed in both life stages in the bryophyte *Physcomitrium patens* (Ortiz-Ramírez et al. 2016). In ferns where both life stages can be independent, even more genes are expressed at both life-stages, e.g., a 97.7% overlap in expression was found in the fern *Polypodium amorphum* (Sigel et al. 2018). Note that different studies might have used different sequencing technologies and methods for quantifying gene expression preventing a formal quantitative comparison; however, these studies collectively suggest a greater potential for haploid selection in plant gametophytes compared to animals, where gene expression in gametes is much more limited (Joseph and Kirkpatrick 2004).

Selection in the haploid gametophytic phase of the lifecycle can be substantial (Immler 2019). The efficacy of both purifying and positive selection is predicted to be higher in the haploid stage due to the unmasking effects in the lack of dominance, e.g., recessive alleles are fully exposed to selection (Haldane 1932; Crow and Kimura 1965; Kondrashov and Crow 1991; Gerstein and Otto 2009). Efficient purifying selection in male gametophytes has been suggested to slow down Y chromosome degeneration (Chibalina and Filatov 2011; Sandler et al. 2018). Several studies have compared the strength of selection on genes expressed exclusively in the haploid or diploid stages. They found stronger purifying and positive selection on pollen-specific genes in the angiosperm *Capsella grandiflora* (Arunkumar et al. 2013) and stronger purifying selection on female gametophyte-specific genes in the gymnosperm *Pinus sylvestris* (Cervantes et al. 2023). However, the signal of differential selection in male haploids can be confounded by the effects of male gametic competition (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et

al. 2022). It is worth noting that the mating system also plays an important role, e.g., in *Arabidopsis thaliana*, pollen-specific genes may experience relaxed purifying selection during the transition to self-pollination (Harrison et al. 2019). Moreover, gene expression level or breadth may affect selection efficacy more than haploid selection (Szövényi et al. 2011; Gossmann et al. 2014).

## Pleiotropy between gametophytes and sporophytes

As haploid and diploid phases of the plant life cycle share the same genome and show widespread gene expression in both phases, an important question arises: how does selection acting on gametophytes influence sporophytes? Haploid selection has important evolutionary consequences to the diploid organism (Otto et al. 2015; Immler and Otto 2018; Immler 2019). Mutations can have pleiotropic effects across plant life stages that are neutral, synergistic (positive), or antagonistic (negative). When a variant only has phase-specific fitness effects in either gametophytes or sporophytes, selection in one phase is by definition neutral to the other phase. When alleles of a variant have concordant or antagonistic fitness effects between life stages, selection in one phase can have synergistic or antagonistic effects in the other, with consequences for the rate of evolution and the maintenance of genetic variation.

Antagonistic pleiotropy between haploid and diploid life stages can be considered as an intralocus conflict when alleles at a single locus have opposite fitness effects between life stages and are subject to ploidally antagonistic selection. Ploidally antagonistic selection can generate balancing selection and maintain genetic variation under certain conditions, but in other cases, variation may still go to fixation despite being under antagonistic selection (Ewing 1977; Immler et al. 2012; Peters and Weis 2018). Variation is more likely to be maintained when alleles are beneficial in the gametophyte but have recessive deleterious effects in the sporophyte, since recessive alleles are masked in the diploid phase, and thus less exposed to selection (Peters and Weis 2018). Additionally, sex differences in selection promote the maintenance of variation under ploidally antagonistic balancing selection (Immler et al. 2012). Despite the theoretical support, empirical tests for ploidally antagonistic selection are limited.

Many studies have tested whether greater pollination intensity, thus stronger pollen competition, affects diploid offspring fitness (Delph and Havens 1998). Different studies have generated mixed results; some found a positive effect of pollen competition on offspring fitness (Palmer

and Zimmerman 1994; Davis 2004; Labouche et al. 2017), due to successful pollen grains improving offspring performance, while others did not (Lankinen et al. 2009; Field et al. 2012; Pélabon et al. 2016). In some cases, there can be genetic trade-offs between pollen competitive ability and sporophytic fitness (Bernasconi et al. 2004). In *Clarkia unguiculata* and *Collinsia heterophylla*, pollen that grows faster pollen tubes or induces earlier stigma receptivity results in reduced seed set for the female parent (Travers and Mazer 2001; Lankinen and Kiboi 2007). In bryophytes, parent-offspring conflict may arise between the paternal haploid genome in the sporophyte and the maternal gametophyte, as sporophytes rely entirely on maternal gametophytes for nutrition during sexual reproduction (Haig and Wilczek 2006; Johnson and Shaw 2016; Shortlidge et al. 2021). In contrast to ploidally antagonistic selection, several plant-breeding studies show that selection on pollen has positive effects on the fitness of sporophytic offspring (synergistic pleiotropy). For example, cold treatment in pollen increases cold tolerance in the offspring in maize (reviewed in Hormaza and Herrero 1992). While these studies have been insightful, the general genomic prevalence of synergistic or antagonistic pleiotropic effects remains unknown.

## Genomic signals of intralocus conflict

With the common genome shared between plant life stages, how can we quantify the extent and nature of pleiotropy between them genome-wide? Here I mainly focus on antagonistic pleiotropy between life stages. There are many similarities in methods studying sexual conflict and conflict between haploid and diploid life stages. Next, I will introduce the development and recent findings from applying population genomics to the study of sexual conflict and how they may help clarify the study of conflict between plant life stages.

Similar to life-stage conflict, sexual conflict or sexual antagonism occurs when males and females have divergent evolutionary interests and different fitness optima over traits (Lande 1980; Bonduriansky and Chenoweth 2009). Sexual dimorphism, i.e., sex differences in phenotypic traits, is widely observed in both plants and animals (Arnqvist and Rowe 2005; Barrett and Hough 2013). Sexual dimorphism in gene expression has been widely used to study sexual conflict, as sex-biased expression is thought to reflect a history of conflict that has now been resolved (Ellegren and Parsch 2007; Mank 2017a). Gene expression level is a key molecular trait (Liu et al. 2019) that could influence an organism's phenotypes and fitness (e.g.,

Charlesworth 2015). Direct tests of sexual antagonism require measuring fitness, which is often challenging. Investigating selection on gene expression provides an indirect method to test for antagonistic selection between sexes or life stages (Price et al. 2022; Stinchcombe and Kelly 2025).

The potential for any future ongoing conflict is determined by the extent of shared genetic architecture underlying phenotypic traits between sexes or life stages. The genetic architecture of a trait describes how genotypic variation contributes to phenotypic diversity under environmental influences, and can refer to the number of genes, their effects, interactions among genes and the environment, and patterns of pleiotropy (i.e., when a gene affects multiple traits), etc. When both sexes share most of their genetic architecture underlying a trait (i.e., a positive and high intersexual genetic correlation, or $r_{MF}$), there is strong potential for ongoing conflict in the presence of antagonistic selection, and the evolution of sexual dimorphism is constrained (Lande 1980; Poissant et al. 2010). Consistent with this prediction, in *Drosophila*, estimates of $r_{MF}$ for gene expression showed high $r_{MF}$ could constrain independent evolution between sexes (e.g., Grieshop et al. 2025; Melo-Gavin et al. 2025). In contrast, when trait covariances are different between sexes (sex-specific genetic variance-covariance matrix **G**) or when genetic influences in one sex do not affect traits in the other sex the same way (an asymmetrical cross-sex covariance matrix **B**), there is less shared genetic architecture between sexes and less potential for ongoing conflict, sexual dimorphism can evolve (Wyman et al. 2013; Houle and Cheng 2021). Whether genetic covariances between life stages limit or promote phenotypic differences between life stages remains an open question.

Intralocus sexual conflict may contribute to the maintenance of genetic variation under specific conditions when different alleles of a variant are maintained by sexually antagonistic balancing selection (Kidwell et al. 1977). Population genomic tests for balancing selection can be used to indirectly test for genes under sexual conflict (Mank 2017b; Ruzicka et al. 2019). For example, current studies have tested for sexually antagonistic balancing selection using diversity statistics, e.g., nucleotide diversity and Tajima's D (Sayadi et al. 2019), or intersexual genetic differentiation, e.g., intersexual $F$st (Cheng and Kirkpatrick 2016; Kasimatis et al. 2021; Ming et al. 2025) or both (Wright et al. 2018; Wright et al. 2019). For example, compared to highly sex-biased genes which represent resolved conflict, unbiased or weakly sex-biased genes show elevated genetic diversity indicative of balancing selection, potentially reflective of ongoing

conflict (Wright et al. 2018; Sayadi et al. 2019). In addition to descriptive statistics, model-based tests will be more powerful in detecting sex-differential selection or antagonistic balancing selection (Cheng and DeGiorgio 2019; Cheng and DeGiorgio 2022; Cole et al. 2024).

However, the conditions for generating sexually antagonistic balancing selection is limited to single large-effect loci (Kidwell et al. 1977; Fry 2010); recent theory on polygenic selection on quantitative traits show very limited ability to detect balancing selection using genomic data (Flintham 2025; Flintham et al. 2025). Notably, sexual antagonism, even without generating balancing selection, can maintain greater genetic variation through mutation-selection balance (Connallon and Clark 2012; Mullon et al. 2012).

The regulation of gene expression is central to the evolution of phenotypic variation. During transcription, gene expression is controlled by both *cis*-acting variation near a gene and *trans*-acting variation elsewhere in the genome. Identifying regulatory variation, e.g., through genome-wide association mapping, can help examine selective forces on regulatory variation and the maintenance of genetic variation (Josephs et al. 2017). *Cis*-regulation usually has a large effect on expression, and most studies have more power to detect it than *trans*-regulatory variation. Testing whether cis-regulatory variation affects sexes or life stages concordantly is important in understanding the potential and resolution of conflict. Studies have found more *cis*-regulatory variation in unbiased than strongly biased genes between sexes, consistent with sexual conflict (Puixeu et al. 2023; but see Mishra et al. 2024). A study on expression quantitative loci (eQTL) across human tissues has shown relatively limited standing variation with sex-specific effects in gene expression, which suggests a potential for ongoing conflict when under sexually antagonistic selection (Oliva et al. 2020). The extent of shared genetic architecture on expression between the sexes and between the diploid and haploid phase of the life cycle remains very poorly known in plants.

## Thesis outline

In my thesis, I investigated the evolutionary genomics of selection and potential conflict between plant life stages in two plant species. Most angiosperms are hermaphrodites with bisexual flowers, while dioecy with separate female and male individuals is less common. My first study system *Rumex hastatulus* (Polygonaceae) is a dioecious, outcrossing, and annual angiosperm with an X/Y sex determination system. Sexual dimorphism in phenotypic and life history traits

of this species is likely associated with differential reproductive roles between sexes during wind pollination (Puixeu et al. 2019). Additionally, the high levels of pollen competition in wind pollinated systems provides an opportunity to test the genetic variation for pollen competition (Friedman and Barrett 2009; Field et al. 2012). Female biased sex ratios are observed in *R. hastatulus*, likely due to a preference of female-determining pollen following Y degeneration (Conn and Blum 1981; Field et al. 2012); gametophytic selection is suggested to play a role on sex chromosome evolution in this species (Hough et al. 2014; Crowson et al. 2017; Sandler et al. 2018). In bryophytes, sex determination occurs during the haploid stage, sexual systems with separate sexes in gametophytes (dioicy) are very common. My second study system *Ceratodon purpureus* (Ditrichaceae) is a dioicous moss with separate sexes and a U/V sex determination system (Carey et al. 2021a). In *C. purpureus*, sexual dimorphism in life history traits (McDaniel 2005; Slate et al. 2017) and the production of volatile organic compounds (Rosenstiel et al. 2012; Kollar et al. 2021) indicate a potential for sexual conflict. In addition to sexual conflict, parent-offspring conflict (Johnson and Shaw 2016; Shortlidge et al. 2021) and sex ratio distortion (McDaniel et al. 2007; Norrell et al. 2014) were suggested to potentially shape sex chromosome evolution of this species (Carey et al. 2021b). Together, the two plant species with their distinctive life cycles are promising systems to study patterns of conflict between haploid and diploid life stages.

In **Chapter 2**, I investigated the potential for conflict between sexes and life stages in *R. hastatulus* by examining the genetic architecture of gene expression in pollen and leaves. Using eQTL mapping, I identified *cis*-regulatory variation and compared the extent to which the variation was shared between sexes and life stages and tested the correlation of eQTL effects on expression. Additionally, I examined selective pressures on cis-regulatory variation by analyzing distribution of allele frequencies. A version of this chapter is currently under review at Molecular Biology and Evolution as: *Yuan M, Sacchi BM, Choudhury BI, Barrett SC, Stinchcombe JR, Wright SI. Cis-regulation of gene expression between sexes and life stages in Rumex hastatulus. https://doi.org/10.1101/2025.06.16.659834*. All authors conceptualized the study and wrote the manuscript. Yuan M and Sacchi BM analyzed the data, Choudhury BI and Yuan M performed the experiment, Stinchcombe JR and Wright SI supervised the study. I am currently revising the chapter following peer review.

In **Chapter 3**, I tested the population genomic signals of intralocus conflict between life stages in *R. hastatulus* and *C. purpureus* and compared the patterns between two species. I examined differential gene expression and genetic diversity patterns to test for signals of intralocus conflict and performed genome-wide scan for balancing selection. I initiated the collaboration with Dr. Stuart McDaniel and his lab on this study, who provided the raw DNA and RNA sequencing data of *C. purpureus*. A version of this chapter has been published at Genome Biology and Evolution as: *Yuan M, Kollar LM, Sacchi BM, Carey SB, Choudhury BI, Jones T, Grimwood J, Barrett SCH, McDaniel SF, Wright SI, Stinchcombe JR. 2025. Testing for the genomic footprint of conflict between life stages in an angiosperm and a moss species. doi.org/10.1093/gbe/evaf138.* All authors conceptualized the study and wrote the manuscript. Yuan M and Sacchi BM analyzed the data; Kollar LM, Choudhury BI, Carey SB, Jones T, Grimwood J and Yuan M performed the experiment; McDaniel SF, Wright SI and Stinchcombe JR supervised the study.

In **Chapter 4**, I combined experimental crosses and pooled sequencing to identify genetic variation for pollen meiotic drive and pollen competition in *R. hastatulus*. I scanned for SNPs showing allele frequency distortion in mature pollen grains (i.e., due to meiotic drive) and developing seeds (i.e., due to pollen competition) and tested sex ratio bias in pollen grains using genomic coverage. This chapter was written in collaboration with John R. Stinchcombe and Stephen I. Wright. All authors conceptualized the study, I performed the experiment and data analysis.

# References

Arnqvist G, Rowe L. 2005. Sexual conflict. Princeton University Press

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol. Biol. Evol.* 30:2475–2486.

Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano Kenji, Takano T, Yano Kentaro, Matsuoka M. 2015. De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, Ljtrans DB. *Plant Cell Physiol.* 56:e5–e5.

Barrett SCH, Hough J. 2013. Sexual dimorphism in flowering plants. *J. Exp. Bot.* 64:67–82.

Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI. 2020. Evolutionary genomics of plant gametophytic selection. *Plant Commun.* 1:100115.

Bernasconi G, Ashman T-L, Birkhead TR, Bishop JDD, Grossniklaus U, Kubli E, Marshall DL,

Schmid B, Skogsmyr I, Snook RR, et al. 2004. Evolutionary ecology of the prezygotic stage. *Science* 303:971–975.

Bonduriansky R, Chenoweth SF. 2009. Intralocus sexual conflict. *Trends Ecol. Evol.* 24:280–288.

Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective. *J. Exp. Bot.* 60:1465–1478.

Carey Sarah B, Jenkins Jerry, Lovell John T., Maumus Florian, Sreedasyam Avinash, Payton Adam C., Shu Shengqiang, Tiley George P., Fernandez-Pozo Noe, Healey Adam, et al. 2021. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Sci. Adv.* 7:eabh2488.

Carey Sarah B., Kollar LM, Mcdaniel SF. 2021. Does degeneration or genetic conflict shape gene content on UV sex chromosomes? *Bryophyte Divers. Evol.* 43:133–149.

Cervantes S, Kesälahti R, Kumpula TA, Mattila TM, Helanterä H, Pyhäjärvi T. 2023. Strong purifying selection in haploid tissue–specific genes of scots pine supports the masking theory. *Mol. Biol. Evol.* 40:msad183.

Charlesworth B. 2015. Causes of natural variation in fitness: Evidence from studies of *Drosophila populations*. *Proc. Natl. Acad. Sci.* 112:1662–1669.

Cheng C, Kirkpatrick M. 2016. Sex-specific selection and sex-biased gene expression in humans and flies.Nachman MW, editor. *PLOS Genet.* 12:e1006170.

Cheng X, DeGiorgio M. 2019. Detection of shared balancing selection in the absence of trans-species polymorphism. *Mol. Biol. Evol.* 36:177–199.

Cheng X, DeGiorgio M. 2022. BalLeRMix+: mixture model approaches for robust joint identification of both positive selection and long-term balancing selection. *Bioinformatics* 38:861–863.

Chettoor AM, Givan SA, Cole RA, Coker CT, Unger-Wallace E, Vejlupkova Z, Vollbrecht E, Fowler JE, Evans MM. 2014. Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. *Genome Biol.* 15:1–23.

Chibalina MV, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* 21:1475–1479.

Cole JM, Scott CB, Johnson MM, Golightly PR, Carlson J, Ming MJ, Harpak A, Kirkpatrick M. 2024. The battle of the sexes in humans is highly polygenic. *Proc. Natl. Acad. Sci.* 121:e2412315121.

Conn JS, Blum U. 1981. Sex ratio of *Rumex hastatulus*: the effect of environmental factors and certation. *Evolution* 35:1108–1116.

Connallon T, Clark AG. 2012. A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics* 190:1477–1489.

Crow JF, Kimura M. 1965. Evolution in sexual and asexual populations. *Am. Nat.* 99:439–450.

Crowson D, Barrett SCH, Wright SI. 2017. Purifying and positive selection influence patterns of gene loss and gene expression in the evolution of a plant sex chromosome system. *Mol. Biol. Evol.* 34:1140–1154.

Darwin C. 1876. The effects of cross and self fertilisation in the vegetable kingdom. J. Murray, London, UK.

Davis SL. 2004. Natural levels of pollination intensity and effects of pollen loads on offspring quality in females of Thalictrum pubescens (Ranunculaceae). *Plant Syst. Evol.* 244:45–54.

Delph LF, Havens K. 1998. Pollen competition in flowering plants. In: Sperm competition and sexual selection. Elsevier. p. 149–173.

Donoghue PCJ, Harrison CJ, Paps J, Schneider H. 2021. The evolutionary emergence of land plants. *Curr. Biol.* 31:R1281–R1298.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* 8:689–698.

Ewing EP. 1977. Selection at the haploid and diploid phases: cyclical variation. *Genetics* 87:195–207.

Field DL, Pickup M, Barrett SCH. 2012. The influence of pollination intensity on fertilization success, progeny sex ratio, and fitness in a wind-pollinated, dioecious plant. *Int. J. Plant Sci.* 173:184–191.

Flintham E. 2025. The evolution of sex-specific gene expression in polygenic traits. *J. Evol. Biol.*:voaf050.

Flintham E, Savolainen V, Otto SP, Reuter M, Mullon C. 2025. The maintenance of genetic polymorphism underlying sexually antagonistic traits. *Evol. Lett.* 9:259–272.

Friedman J, Barrett SC. 2009. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Ann. Bot.* 103:1515–1527.

Fry JD. 2010. The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution* 64:1510–1516.

Gerstein AC, Otto SP. 2009. Ploidy and the causes of genomic evolution. *J. Hered.* 100:571–581.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 31:574–583.

Grieshop K, Liu MJ, Frost RS, Lindsay MP, Bayoumi M, Brengdahl MI, Molnar RI, Agrawal AF. 2025. Expression divergence in response to sex-biased selection. *Mol. Biol. Evol.* 42:msaf099.

Gutiérrez-Valencia J, Fracassetti M, Horvath R, Laenen B, Désamore A, Drouzas AD, Friberg M, Kolář F, Slotte T. 2022. Genomic signatures of sexual selection on pollen-expressed genes in *Arabis alpina*.de Meaux J, editor. *Mol. Biol. Evol.* 39:msab349.

Haig D, Wilczek A. 2006. Sexual conflict and the alternation of haploid and diploid generations. *Philos. Trans. R. Soc. B Biol. Sci.* 361:335–343.

Haldane JBS. 1932. The causes of evolution. Reprinted 1990. Princeton Univ. Press, Princeton, NJ.

Harris BJ, Clark JW, Schrempf D, Szöllősi GJ, Donoghue PCJ, Hetherington AM, Williams TA.

2022. Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nat. Ecol. Evol.* 6:1634–1643.

Harrison CJ. 2017. Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150490.

Harrison MC, Mallon EB, Twell D, Hammond RL. 2019. Deleterious mutation accumulation in *Arabidopsis thaliana* pollen genes: a role for a recent relaxation of selection. *Genome Biol. Evol.* 11:1939–1951.

Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* 132:640–652.

Hormaza JI, Herrero M. 1992. Pollen selection. *Theor. Appl. Genet.* 83:663–672.

Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. 2014. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proc. Natl. Acad. Sci.* 111:7713–7718.

Houle D, Cheng C. 2021. Predicting the evolution of sexual dimorphism in gene expression. *Mol. Biol. Evol.* 38:1847–1859.

Immler S. 2019. Haploid selection in "diploid" organisms. *Annu. Rev. Ecol. Evol. Syst.* 50:219–236.

Immler S, Arnqvist G, Otto SP. 2012. Ploidally antagonistic selection maintains stable genetic polymorphism: ploidally antagonistic selection. *Evolution* 66:55–65.

Immler S, Otto SP. 2018. The evolutionary consequences of selection at the haploid gametic stage. *Am. Nat.* 192:241–249.

Johnson MG, Shaw AJ. 2016. The effects of quantitative fecundity in the haploid stage on reproductive success and diploid fitness in the aquatic peat moss Sphagnum macrophyllum. *Heredity* 116:523–530.

Joseph S, Kirkpatrick M. 2004. Haploid selection in animals. *Trends Ecol. Evol.* 19:592–597.

Josephs EB, Stinchcombe JR, Wright SI. 2017. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol.* 214:21–33.

Kasimatis KR, Abraham A, Ralph PL, Kern AD, Capra JA, Phillips PC. 2021. Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics* 217:iyaa015.

Kidwell JF, Clegg MT, Stewart FM, Prout T. 1977. Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics* 85:171–183.

Kollar LM, Kiel S, James AJ, Carnley CT, Scola DN, Clark TN, Khanal T, Rosenstiel TN, Gall ET, Grieshop K, et al. 2021. The genetic architecture of sexual dimorphism in the moss *Ceratodon purpureus*. *Proc. R. Soc. B Biol. Sci.* 288:20202908.

Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? *Nature* 351:314–315.

Labouche A-M, Richards SA, Pannell JR. 2017. Effects of pollination intensity on offspring number and quality in a wind-pollinated herb. *J. Ecol.* 105:197–208.

Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters.

*Evolution* 34:292.

Lankinen A, Kiboi S. 2007. Pollen donor identity affects timing of stigma receptivity in *Collinsia heterophylla* (Plantaginaceae): a sexual conflict during pollen competition? *Am. Nat.* 170:854–863.

Lankinen Å, Maad J, Armbruster WS. 2009. Pollen-tube growth rates in *Collinsia heterophylla* (Plantaginaceae): one-donor crosses reveal heritability but no effect on sporophytic-offspring fitness. *Ann. Bot.* 103:941–950.

Lewis LA, McCourt RM. 2004. Green algae and the origin of land plants. *Am. J. Bot.* 91:1535–1556.

Liu X, Li YI, Pritchard JK. 2019. Trans effects on gene expression can drive omnigenic inheritance. *Cell* 177:1022-1034.e6.

Lloyd DG. 1974. Female-predominant sex ratios in angiosperms. *Heredity* 32:35–44.

Mable BK, Otto SP. 1998. The evolution of life cycles with haploid and diploid phases. *BioEssays* 20:453–462.

Mank JE. 2017a. The transcriptional architecture of phenotypic dimorphism. *Nat. Ecol. Evol.* 1:0006.

Mank JE. 2017b. Population genetics of sexual conflict in the genomic era. *Nat. Rev. Genet.* 18:721–730.

McDaniel SF. 2005. Genetic correlations do not constrain the evolution of sexual dimorphism in the moss Ceratodon purpureus. *Evol. Int. J. Org. Evol.* 59:2353–2361.

McDaniel SF. 2021. Bryophytes are not early diverging land plants. *New Phytol.* 230:1300–1304.

McDaniel SF, Willis JH, Shaw AJ. 2007. A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. *Genetics* 176:2489–2500.

Melo-Gavin C, Grieshop K, Agrawal AF. 2025. Genomic response to sex-separated gene pools. :2025.07.18.665591. Available from: https://www.biorxiv.org/content/10.1101/2025.07.18.665591v1

Ming MJ, Cheng C, Kirkpatrick M, Harpak A. 2025. No evidence for sex-differential transcriptomes driving genome-wide sex-differential natural selection. *Am. J. Hum. Genet.* 112:254–260.

Mishra P, Barrera TS, Grieshop K, Agrawal AF. 2024. *Cis*-regulatory variation in relation to sex and sexual dimorphism in *Drosophila melanogaster*. *Genome Biol. Evol.* 16:evae234.

Mulcahy DL, Mulcahy GB. 1987. The effects of pollen competition. *Am. Sci.* 75:44–50.

Mullon C, Pomiankowski A, Reuter M. 2012. The effects of selection and genetic drift on the genomic distribution of sexually antagonistic alleles. *Evolution* 66:3743–3753.

Norrell TE, Jones KS, Payton AC, McDaniel SF. 2014. Meiotic sex ratio variation in natural populations of Ceratodon purpureus (Ditrichaceae). *Am. J. Bot.* 101:1572–1576.

Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, Parsana P, Kasela S, Balliu B, Viñuela A, et al. 2020. The impact of sex on gene expression across

human tissues. *Science* [Internet] 369. Available from: https://science.sciencemag.org/content/369/6509/eaba3066

Ortiz-Ramírez C, Hernandez-Coronado M, Thamm A, Catarino B, Wang M, Dolan L, Feijó JA, Becker JD. 2016. A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Mol. Plant* 9:205–220.

Otto SP, Scott MF, Immler S. 2015. Evolution of haploid selection in predominantly diploid organisms. *Proc. Natl. Acad. Sci.* 112:15952–15957.

Palmer TM, Zimmerman M. 1994. Pollen competition and sporophyte fitness in Brassica campestris: does intense pollen competition result in individuals with better pollen? *Oikos*:80–86.

Pélabon C, Hennet L, Bolstad GH, Albertsen E, Opedal ØH, Ekrem RK, Armbruster WS. 2016. Does stronger pollen competition improve offspring fitness when pollen load does not vary? *Am. J. Bot.* 103:522–531.

Peters MAE, Weis AE. 2018. Selection for pollen competitive ability in mixed-mating systems. *Evolution* 72:2513–2536.

Poissant J, Wilson AJ, Coltman DW. 2010. Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution* 64:97–107.

Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022. Detecting signatures of selection on gene expression. *Nat. Ecol. Evol.* 6:1035–1045.

Puixeu G, Macon A, Vicoso B. 2023. Sex-specific estimation of *cis* and *trans* regulation of gene expression in heads and gonads of *Drosophila melanogaster*. *G3 GenesGenomesGenetics* 13:jkad121.

Puixeu G, Pickup M, Field DL, Barrett SCH. 2019. Variation in sexual dimorphism in a wind-pollinated plant: the influence of geographical context and life-cycle dynamics. *New Phytol.* 224:1108–1120.

Qiu Y, Taylor AB, McManus HA. 2012. Evolution of the life cycle in land plants. *J. Syst. Evol.* 50:171–194.

Rosenstiel TN, Shortlidge EE, Melnychenko AN, Pankow JF, Eppley SM. 2012. Sex-specific volatile compounds influence microarthropod-mediated fertilization of moss. *Nature* 489:431–433.

Rutley N, Twell D. 2015. A decade of pollen transcriptomics. *Plant Reprod.* 28:73–89.

Ruzicka F, Hill MS, Pennell TM, Flis I, Ingleby FC, Mott R, Fowler K, Morrow EH, Reuter M. 2019. Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. *PLoS Biol.*

Sandler G, Beaudry FEG, Barrett SCH, Wright SI. 2018. The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants. *Evol. Lett.* 2:368–377.

Sayadi A, Martinez Barrio A, Immonen E, Dainat J, Berger D, Tellgren-Roth C, Nystedt B, Arnqvist G. 2019. The genomic footprint of sexual conflict. *Nat. Ecol. Evol.* 3:1725–1730.

Shortlidge EE, Carey SB, Payton AC, McDaniel SF, Rosenstiel TN, Eppley SM. 2021. Microarthropod contributions to fitness variation in the common moss *Ceratodon purpureus*. *Proc. R. Soc. B Biol. Sci.* 288:20210119.

Sigel EM, Schuettpelz E, Pryer KM, Der JP. 2018. Overlapping patterns of gene expression between gametophyte and sporophyte phases in the fern *Polypodium amorphum* (Polypodiales). *Front. Plant Sci.* 9:1450.

Slate ML, Rosenstiel TN, Eppley SM. 2017. Sex-specific morphological and physiological differences in the moss *Ceratodon purpureus* (Dicranales). *Ann. Bot.* 120:845–854.

Stinchcombe JR, Kelly JK. 2025. Measuring natural selection on the transcriptome. *New Phytol.*

Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ. 2011. Generation-biased gene expression in a bryophyte model system. *Mol. Biol. Evol.* 28:803–812.

Taylor TN, Kerp H, Hass H. 2005. Life history biology of early land plants: deciphering the gametophyte phase. *Proc. Natl. Acad. Sci.* 102:5892–5897.

Travers SE, Mazer SJ. 2001. Trade-offs between male and female reproduction associated with allozyme variation in phosphoglucoisomerase in an annual plant (*Clarkia unguiculata*: Onagraceae). *Evolution* 55:2421–2428.

Walbot V, Evans MMS. 2003. Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* 4:369–379.

Walsh NE, Charlesworth D. 1992. Evolutionary Interpretations of Differences in Pollen Tube Growth Rates. *Q. Rev. Biol.* 67:19–37.

Williams JH, Mazer SJ. 2016. Pollen—tiny and ephemeral but not forgotten: new ideas on their ecology and evolution. *Am. J. Bot.* 103:365–374.

Wright AE, Fumagalli M, Cooney CR, Bloch NI, Vieira FG, Buechel SD, Kolm N, Mank JE. 2018. Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evol. Lett.* 2:52–61.

Wright AE, Rogers TF, Fumagalli M, Cooney CR, Mank JE. 2019. Phenotypic sexual dimorphism is associated with genomic signatures of resolved sexual conflict. *Mol. Ecol.* 28:2860–2871.

Wyman MJ, Stinchcombe JR, Rowe L. 2013. A multivariate view of the evolution of sexual dimorphism. *J. Evol. Biol.* 26:2070–2080.

# Chapter 2: *Cis*-regulation of gene expression between sexes and life stages in *Rumex hastatulus*

*A version of this chapter is currently under review at Molecular Biology and Evolution as:*

*Yuan M, Sacchi BM, Choudhury BI, Barrett SC, Stinchcombe JR, Wright SI. Cis-regulation of gene expression between sexes and life stages in Rumex hastatulus. https://doi.org/10.1101/2025.06.16.659834. All authors conceptualized the study and wrote the manuscript. Yuan M and Sacchi BM analyzed the data, Choudhury BI and Yuan M performed the experiment, Stinchcombe JR and Wright SI supervised the study.*

## Abstract

The potential for conflict between sexes and life stages while sharing predominantly the same genome has important evolutionary consequences. In dioecious angiosperms, genes beneficial for the haploid pollen stage may reduce the fitness of diploid offspring of both males and females. Such conflict between sexes and life stages can in some cases maintain genetic variation. However, we still lack understanding of the extent of shared genetic architecture for gene expression between the sexes or life stages in plants, a key component for predicting the scope for conflict. We performed expression quantitative trait loci (eQTL) mapping to test if standing variation affects sexes and life stages differently using a population sample of the dioecious outcrossing plant *Rumex hastatulus*. We compared effect sizes and allele frequencies of *cis*-eQTLs in male and female leaf tissues and pollen and tested for genotype-by-sex interactions for gene expression. We found stronger shared genetic architecture between sexes than between the life stages. In addition, comparisons of the site frequency spectra for *cis*-eQTLs to a null distribution found no evidence for a genome-wide pattern consistent with purifying selection. Our results suggest that any conflict over optimal gene expression between pollen and leaves may be easily resolved due to their distinct genetic architectures, whereas there is more scope for conflict between the sexes in gene expression for leaves. Our study highlights the use of eQTLs to infer the scope of shared genetic architecture and for investigating the evolution of conflict between sexes and life stages in dioecious species.

# Introduction

Evolutionary conflict may arise when different sexes or life-cycle phases of an organism have different fitness optima while sharing a common genome (Ewing 1977; Arnqvist and Rowe 2005; Haig and Wilczek 2006; Schärer et al. 2015). For example, sexual conflict occurs when males and females have different fitness optima involving reproduction and this can drive sexual dimorphism in phenotypic traits (reviewed in Arnqvist and Rowe 2005). In plants, the diploid and haploid phases of the life cycle (hereafter 'life stages') are distinct in their morphology, development, and genetics despite sharing extensive overlap in gene expression, creating potential for genetic trade-offs (Haig and Wilczek 2006; Beaudry et al. 2020). Both sexual and life-stage conflict have been suggested to generate balancing selection and stable genetic variation under some conditions (Kidwell et al. 1977; Immler et al. 2012). However, we lack comprehensive understanding of the traits and genetic architecture underlying the conflict between sexes or life stages. Here, we use expression quantitative trait locus (eQTL) mapping to identify genome-wide *cis*-regulatory variation and investigate the extent of shared genetic architecture between sexes and life stages and the scope for sexual and life-stage conflict in a dioecious plant species.

Current empirical examples of life-stage conflict often involve both sexes (i.e., a conflict both between the sexes and life stages). For example, alleles beneficial for haploid gametophytic competition were shown to reduce the fitness of female diploid offspring in the hermaphroditic annuals *Clarkia* (Travers and Mazer 2001) and *Collinsia* (Lankinen and Kiboi 2007). In *C. heterophylla*, pollen donors that induce early stigma receptivity, thus increasing paternity, reduced seed set (Lankinen and Kiboi 2007; Madjidian and Lankinen 2009; Lankinen et al. 2017). However, sexual selection and sexual conflict in sessile plants involving one or both life stages still remain understudied compared to the extensive evidence from numerous animal species (Prasad and Bedhomme 2006; Moore and Pannell 2011). The diverse reproductive systems in plants provide a unique opportunity to study sexual conflict, from conflict between female and male function in hermaphrodites (Duffy et al. 2021) to sexual dimorphisms in dioecious species and the resolution of intralocus sexual conflict (Delph et al. 2010). Additionally, because of the considerable scope for both sexual and life-stage conflict in plants, whether there is more scope for one type of conflict than the other, or whether one type of conflict is more resolved than the other, remains unanswered.

The scope for conflict will depend on the degree of shared genetic architecture between sexes or life stages. For example, high genetic correlation in a trait between sexes may constrain the responses to antagonistic selection and the evolution of sexual dimorphism (Lande 1980; Poissant et al. 2010; Wyman et al. 2013; Grieshop et al. 2025). Gene expression is an important molecular trait underlying phenotypic variation, and sexual dimorphism in gene expression is likely a key driver of the phenotypic divergence between sexes (Mank 2017). Genetic covariances between sexes have been inferred through gene expression studies in *Drosophila melanogaster* suggesting considerable scope for ongoing conflict (Innocenti and Morrow 2010; Griffin et al. 2013; Houle and Cheng 2021). Similarly, whether genetic correlations in traits across life stages commonly constrain their independent evolution under opposing selection is an important unresolved question.

Expression quantitative trait loci (eQTL) mapping is useful for identifying loci underlying regulatory variation and testing the genome-wide importance of selection (Kudaravalli et al. 2009; Lee et al. 2014; Josephs et al. 2015). Genetic variation in the regulatory region near a gene, i.e., *cis*-regulatory variation, has a relatively large effect on gene expression, making it easier to map *cis*-eQTLs in association studies (Josephs et al. 2017; Signor and Nuzhdin 2018). Selective pressure on eQTLs can be inferred from their distribution of allele frequencies; an excess of rare alleles will indicate purifying selection that removes deleterious mutations, whereas an excess of common alleles indicates balancing selection maintaining genetic variation (or weaker purifying selection). In the leaf tissues of *Capsella grandiflora*, trans-eQTLs were found to be under stronger purifying selection than *cis*-eQTLs (Josephs et al. 2015; Josephs et al. 2020). In contrast, in floral tissues of *Mimulus guttatus*, trans-eQTLs showed signatures of purifying selection whereas *cis*-eQTLs showed signatures of 'balancing selection' based on their allele frequencies (Brown and Kelly 2021). These studies from two species suggest eQTLs can undergo different types of selection, suggesting the possibility of differential selection on genes expressed in different sexes or life stages.

In addition to testing the genome-wide importance of purifying versus balancing selection on regulatory variation, eQTLs can be used to test whether standing variation affects sexes or life stages concordantly. A study on *cis*-eQTLs in 44 human tissues between sexes found that despite a moderate number of genes showing sex-biased expression, *cis*-eQTLs showing Genotype × Sex interaction were much rarer (Oliva et al. 2020). In *Drosophila melanogaster* somatic tissues,

more *cis*-regulatory variation was found in unbiased and moderately biased than strongly biased genes between sexes. This finding suggested that the former genes experience potential ongoing sexual conflict in regulatory variation (Puixeu et al. 2023; but see Mishra et al. 2024). Despite the importance of understanding the scope of genome-wide conflict, we still lack tests on the extent of shared genetic architecture between sexes or life stages, especially in dioecious plant species where both types of conflict in one system can be compared.

In this study, we tested *cis*-regulation in gene expression through *cis*-eQTL mapping in the dioecious, wind-pollinated and obligately outcrossing annual *Rumex hastatulus* (Polygonaceae). Female biased sex ratios are observed in *R. hastatulus*, likely due to greater competitive ability of female-determining pollen following Y-degeneration (Field et al. 2013; Pickup and Barrett 2013), and gametophytic selection is suggested to play a role in sex chromosome evolution in this species (Hough et al. 2014; Crowson et al. 2017; Sandler et al. 2018). Several observations in *R. hastatulus* indicate the potential for antagonism between sex and life stages, especially during pollen competition. Sexual dimorphism related to wind pollination in both vegetative and reproductive traits occurs in *R. hastatulus*, e.g., males are taller and produce more inflorescences during peak flowering, presumably to maximize access to female flowers (Puixeu et al. 2019). The potential conflict between sexes and life stages in *R. hastatulus* makes it a promising experimental system to test the direction of selection on gene expression between sexes and life stages and also to compare the scopes for sexual versus life-stage conflict.

We tested differential gene expression between sexes and life stages, and mapped *cis*-eQTLs separately in leaf and pollen tissues to compare their effect sizes and allele frequencies. We found significantly fewer differentially expressed genes between sexes than between life stages, and most of the highly sex-biased genes were on the sex chromosomes. There was a greater overlap in eGenes (i.e., genes with their expression level affected by eQTLs) and eQTLs between sexes than life stages. Effect sizes of eQTLs were positively correlated between sexes and life stages, with a much stronger correlation between sexes, indicating more shared genetic architecture between sexes. Consistent with this result, we found limited Genotype × Sex interaction in gene expression. There were a large number of eQTLs driving the divergent expression between life stages. Lastly, we found similar allele frequency distributions of eQTLs among tissues, suggesting similar selective pressures in eQTLs between sexes and life stages but no excess of rare alleles among eQTLs in any tissues when compared to a null distribution.

# Results

## *Differential expression between sexes and life stages*

We generated leaf DNA and RNA sequences and pollen RNA sequences from 78 male-female sibling pairs in a single large population (see Methods). We used male and female leaf to represent the sexes and used male leaf and pollen to represent diploid and haploid life stages, respectively. We performed differential gene expression analyses to identify sex-biased genes in leaves and life stage-biased genes between pollen and male leaf tissues (Supplementary Table A3). Among the 37,659 annotated genes across the genome, we first filtered for genes with evidence of expression in at least one of the two tissues being compared, resulting in 19,494 genes in the between-sex comparison and 21,337 genes in the between-life-stage comparison. In a principal component analysis (PCA) of RNA read counts, pollen and male leaf samples were separated by PC1 that explained 99% variance both before and after excluding sex-specific regions on sex chromosomes (Supplementary Figure A1). Male and female leaf samples were separated by PC1 which explained 52% of the variance when sex-specific regions were included but were not separated based on autosomal and pseudoautosomal (PAR) genes (Supplementary Figure A1). Consistent with the PCA results, there were over 10-fold more differentially expressed (DE) genes between life stages (16,649) than between sexes (1,496) (adjusted $p <$ 0.05, fold change > 2, Supplementary Table A3).

We tested for enrichment of DE genes on sex chromosomes and the overlap between sex-biased and life stage-biased genes. Sex-biased genes were highly enriched on sex chromosomes (Supplementary Table A3), with 378 out of 405 female-biased genes on the X-specific regions and 1,026 out of 1,091 male-biased genes on the Y-specific regions (adjusted $p < 0.05$, fold change > 2). Among 6,702 pollen-biased genes, there was an excess of over-expression of these genes in male leaf compared to female leaf (30 on autosomes and PAR, 1 on X, 269 on Y) tissue compared to female-biased genes (7 on autosomes and PAR, 63 on X, 3 on Y) (Fisher's exact test, $p = 0.0028$). Although it was a relatively small subset of pollen-based genes, the larger overlap between male leaf-biased and pollen-biased genes was mostly driven by the Y-specific regions. This finding implies that selection for higher gametophytic expression in pollen could have pleiotropic effects on expression in male leaves (and *vice versa*), which could lead to antagonistic or synergistic pleiotropy for gene expression depending on the direction of selection on expression between life stages.

A previous study in two *Rumex* species found an enrichment of pollen-biased genes on sex chromosomes relative to autosomes, suggesting a role of male gametophytic selection in Y chromosome evolution (Sandler et al. 2018). With a new PacBio genome assembly containing X and Y phased assemblies (Sacchi et al., unpublished manuscript), we attempted to replicate these results with our gene expression data in *R. hastatulus*. In accord with the earlier results, we found a significant enrichment of pollen-biased genes on the Y chromosome compared to the autosomes or the X chromosome (Supplementary Figure A2, Fisher's exact test, $p = 0.007$ for autosomes, $p = 0.001$ for X) specifically for highly pollen-biased genes (fold change > 4). Among moderately pollen-biased genes (fold change > 2), we found a significant enrichment of pollen-biased genes on the Y chromosome only when compared to the X chromosome (Supplementary Figure A2, Fisher's exact test, $p = 0.031$ for X, $p > 0.1$ for autosomes).

## eQTLs between sexes and life stages

To test whether *cis*-eQTLs affect expression concordantly between sexes or life stages, we performed *cis*-eQTL mapping separately in male leaf, female leaf and pollen. For each gene, we tested the nearby SNPs for significant associations between their genotypes and this gene's expression level, significant associations of gene-SNP pairs after false discovery rate corrections would be considered as *cis*-eQTLs (see Methods). For leaf samples, we mapped *cis*-eQTLs using leaf expression and leaf genotypes in males and females separately. For pollen samples, we used pollen expression and male leaf genotypes as they were from the same male individual and shared the same genotypes. We obtained 4,509,904 and 3,288,995 SNPs on autosomes from 74 female and 69 male leaf DNA samples, respectively, after removing variants with low minor allele frequency or strong Hardy–Weinberg deviations. Based on Tracy-Widom tests, we found 5 and 6 significant PCs for female and male samples explaining 6.53% and 7.7% of genetic variation, respectively (TW statistic $\geq 1.068$, $p \leq 0.0441$ for females; TW statistic $\geq 2.288$, $p \leq 0.00634$ for males); we included the significant PCs as covariates in eQTL mapping.

We identified *cis*-eQTLs and genes whose expression was influenced by eQTLs ("eGenes") in male leaf, female leaf and pollen (Table 2.1, Supplementary Figure A3). We found more shared eGenes and eQTLs between male and female leaf than between male leaf and pollen (Figure 2.1). The number of eQTLs per eGene ranged from 1 to 126 in pollen, 1 to 170 in male leaf, and 1 to 268 in female leaf, with a median of 2 eQTLs per eGene in both male leaf and pollen, and 3

in female leaf. We found eQTLs were located closer to their associated eGenes than the nearby SNPs tested in all tissues (median distance to the transcription start site: 7718 - 7893 bp for eQTLs vs. 8982 - 9254 bp for nearby SNPs). There were similar numbers of eQTLs upstream and downstream of the transcription start site (Supplementary Figure A4). We found a strong negative correlation between minor allele frequencies and effect sizes of eQTLs in all tissues when selecting the most significant eQTL (top eQTL) per eGene (Spearman correlation, $r = -0.60 \sim -0.65$, $p < 10^{-16}$) and when selecting a random eQTL per eGene ($r = -0.71 \sim -0.75$, $p < 10^{-16}$) (Supplementary Figure A5). There was a weak but significant negative correlation between effect sizes of eQTLs and their distance to the transcription start site in all tissues when selecting the top eQTL (Spearman correlation, $r = -0.04 \sim -0.09$, $p < 0.041$) and also when selecting a random eQTL ($r = -0.06 \sim -0.11$, $p < 0.001$).

Table 2.1. eQTL mapping in male leaf, female leaf and pollen on autosomes. FDR = 0.1.

| | Number of samples | Number of genes tested | Number of nearby SNPs tested | Number of eGenes | Number of eQTLs | Number of eQTLs affecting multiple genes | Number of secondary eQTLs |
|---|---|---|---|---|---|---|---|
| Male leaf | 69 | 14,777 | 1,272,993 | 2,421 | 16,674 | 486 | 367 |
| Female leaf | 74 | 14,564 | 1,687,298 | 3,442 | 34,238 | 905 | 722 |
| Pollen | 69 | 13,785 | 1,182,102 | 1,475 | 9,456 | 201 | 159 |
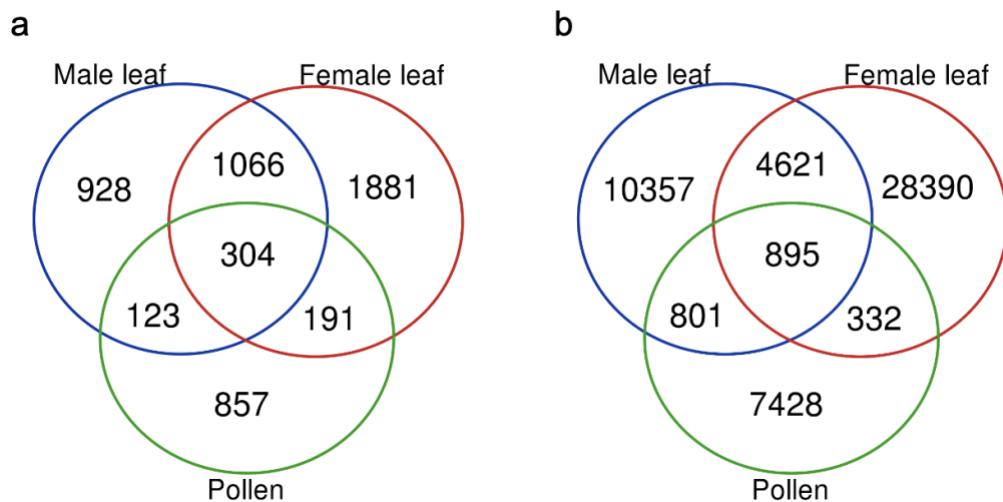
Figure 2.1. Venn diagrams of eGenes (a) and eQTLs (b) between sexes and life stages on autosomes of *Rumex hastatulus*. FDR = 0.1.

In addition to the primary eQTLs in each tissue, for each eGene we examined the independence of its eQTLs and identified secondary eQTLs (Table 2.1, see Methods). Secondary eQTLs were considered as conditionally independent after accounting for the effect of the primary eQTL of each eGene. We found primary eQTLs were significantly closer to the transcription start site than secondary eQTLs in female leaf and male leaf tissue (*t*-test, $p$ = 0.02, 0.009, 0.32 for female leaf, male leaf, pollen, respectively). There were 40 - 43% of secondary eQTLs that had opposite effects compared to the primary eQTL of the eGene, consisting of 0.7 - 0.9 % of all eQTLs in each tissue (158 for male leaf, 291 for female leaf, 63 for pollen). The minor allele frequencies of primary and secondary eQTLs were not significantly different (Mann-Whitney *U*-test, $p$ = 0.12, 0.37, 0.22 for male leaf, female leaf, pollen, respectively).

We performed gene ontology enrichment of eGenes in each tissue and examined whether they showed transcriptional or protein biosynthesis functions. Different tissues showed different patterns of functional enrichment (Supplementary Table A4). For example, female leaf tissue showed enrichment for protein metabolic process (GO:0019538) whereas pollen showed enrichment for mRNA metabolic process (GO:0016071) and negative regulation of gene expression (GO:0010629). Additionally, protein deubiquitination (GO:0016579) was enriched in both male leaf and pollen, and regulation of RNA metabolic process (GO:0051252) was enriched in female leaf and pollen.

*Shared genetic basis between sexes and life stages*

To investigate the extent of shared genetic architecture in expression between sexes and life stages, we first estimated the correlation in expression level across genes between sexes and life stages. For comparing sexes, we estimated the correlation in expression between male and female siblings, which is a phenotypic approximation of the genetic correlation between sexes ($r_{mf}$). For comparing life stages, this represents the phenotypic correlation between traits, which reflects both environmental and genetic variances. We found expression levels were more positively correlated between sexes than life stages based on all genes and eGenes (Figure 2.2, Mann-Whitney *U*-test, $p < 10^{-16}$ in both comparisons). Even though correlations between life stages involve the same individual (as opposed to siblings of the opposite sex, for comparing the

sexes), the stronger correlation in expression between sexes, suggests potentially more shared genetic basis between sexes than life stages.
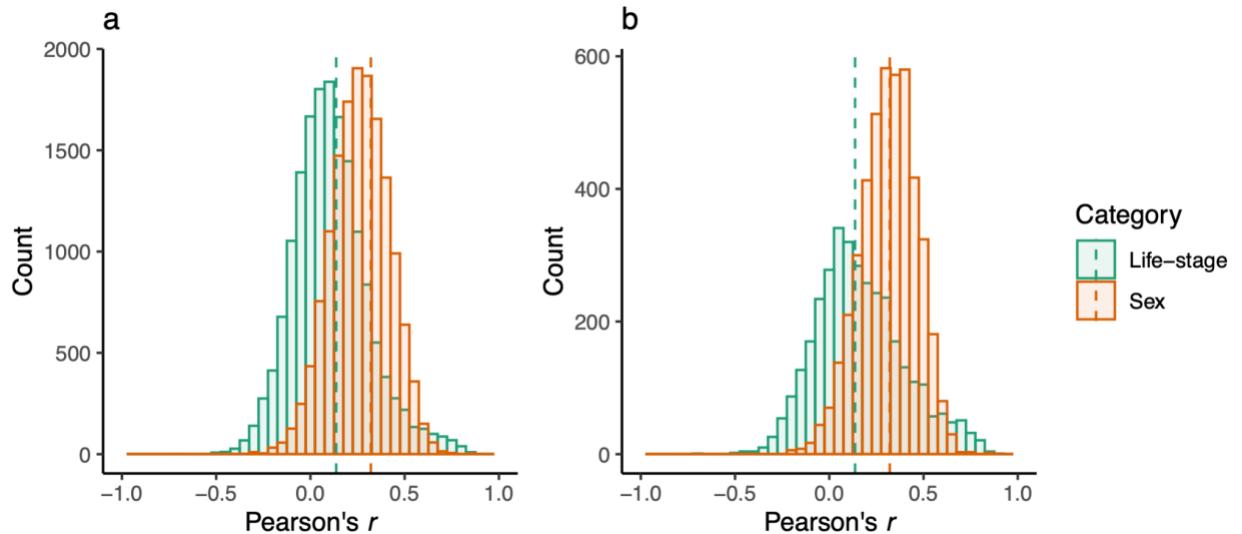


Figure 2.2. Distribution of Pearson correlation coefficients in gene expression between sexes or life stages in all genes (a) and in eGenes (b) in *Rumex hastatulus*. The dashed line represents the median of each group: 0.093 (Sex) 0.27 (Life-stage) in a, and 0.14 (Sex) 0.32 (Life-stage) in b. Numbers of genes: 14992 (Sex) and 16411 (Life-stage) in a, 4493 (Sex) and 3469 (Life-stage) in b.

Next, we tested whether *cis*-regulatory variation had concordant or discordant effects in gene expression between sexes and life stages. We filtered for nearby SNPs tested in both sexes or life stages that were the top eQTL for the eGene in at least one sex or life stage, resulting in 2,420 eQTLs in the sex comparison and 3,077 eQTLs in the life-stage comparison. The majority of eQTLs showed concordant effects between both sexes and life stages, with their effect sizes positively correlated (Figure 2.3). The correlation of eQTL effect sizes was much stronger between sexes (Spearman correlation, $r = 0.72$, $p < 10^{-16}$) than between life stages ($r = 0.35$, $p < 10^{-16}$), suggesting a stronger shared genetic architecture between sexes (Figure 2.3), as would be predicted from the phenotypic correlations in Figure 2.2. There were 170 eQTLs with opposite effects between sexes, all of which were the top eQTL in one sex but not an eQTL in the other sex (light orange and light purple dots in Figure 2.3a). There were 974 eQTLs with opposite effects between life stages, 19 of which were eQTLs in both life stages (Supplementary Figure A6). Functional enrichment of eGenes with eQTLs showing opposite effects between sexes or

life stages showed an enrichment in processes relevant to transcription including negative regulation of gene expression (GO:0010629) and regulation of DNA-templated transcription (GO:0006355), and processes post transcription including regulation of RNA and protein metabolic processes (GO:0051252, GO:0051246) (Supplementary Table A5).
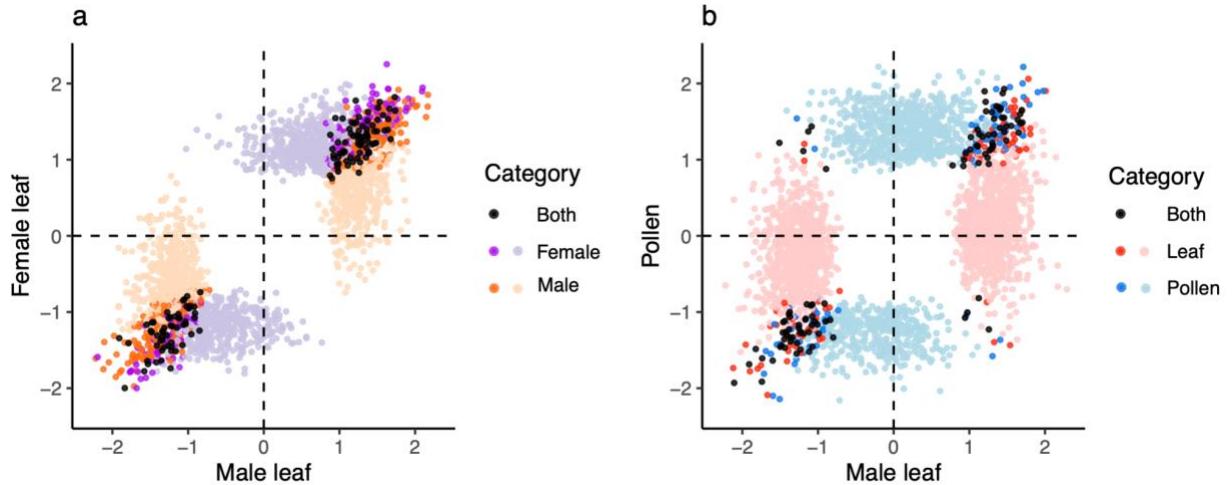


Figure 2.3. Correlation of effect sizes of autosomal SNPs tested between sexes (a) and life stages (b) in *Rumex hastatulus*. Black: top eQTL of the eGene in both sexes or life stages; darker colors: top eQTL in the labelled sex/life stage, significant but not the top eQTL in the other sex/life stage; lighter colors: top eQTL in the labelled sex/life stage but not significant in the other sex/life stage. Spearman correlation coefficients: 0.72 (a), 0.35 (b).

We obtained 3,684,937 SNPs on autosomes from 143 leaf DNA samples to test for Genotype × Sex interactions in *cis*-regulatory variation. We found one significant PC in all leaf samples combined that explained 2% of genetic variation (TW statistic = 1.328, $p = 0.0302$) and we included it as a covariate in eQTL mapping. We tested 14,587 genes and 1,083,706 nearby SNPs for the effect of Genotype, Sex, and Genotype × Sex interaction on expression. After the correction for multiple testing (Davis et al. 2016), we found 5 eQTLs showing significant Genotype × Sex interaction (adjusted $p < 0.1$, see Methods), all of which changed expression of their eGene in the same direction but with different magnitudes (Supplementary Figure A7). The small number of eQTLs showing Genotype × Sex is consistent with the previous result that eQTLs showed strongly correlated effects between sexes (Figure 2.3a).

We tested 16,218 genes and 1,404,801 nearby SNPs for *cis*-eQTLs affecting the degree of expression differences between life stages. Specifically, we calculated the effect sizes of the minor allele of each SNP on the differences in expression levels between male leaf tissue and pollen. We found 9,873 eQTLs affecting the expression differences between life stages in 1,564 genes. There were more eQTLs in which the minor allele increased expression differences (6,173) than reduced them (3,700), suggesting new mutations (typically represented by the minor alleles) more often drive the divergence in expression between life stages than constrain it. We found 1,464 of 1,564 genes whose expression differences between life stages were affected by eQTLs were significant DE genes between pollen and male leaf tissue (adjust $p < 0.05$, no fold change cutoff). Compared to all DE genes, these 1,464 genes were enriched for high expression bias (fold change > 8) and underrepresented for low expression bias (fold change < 2) (Chi-squared test, $p < 10^{-10}$; Pearson residuals = 3.37, -4.95, respectively), likely reflecting that highly pollen-biased genes have more pollen-specific regulatory functions. We found rRNA processing (GO:0006364) and regulation of RNA metabolic processes (GO:0051252) were enriched in genes with eQTLs affecting life-stage expression differences (Supplementary Table A6). Lastly, we identified 141 conditionally independent eQTLs affecting life-stage expression differences, 54 of which had opposite effects compared to the primary eQTL of the same gene.

We examined whether eQTLs for expression differences between life stages were also identified as eQTLs in male leaf tissue or pollen separately. We mapped eQTLs in male leaf and pollen testing the effects of minor alleles on expression. There were 1,051 genes whose expression differences were significantly affected by *cis*-regulatory variation and were also tested for eQTLs in both male leaf and pollen individually. Of the 5,946 eQTLs associated with expression differences in these 1,051 genes, 5,027 were also identified as eQTLs in at least one life stage, while the remaining did not pass the multiple testing filters to be eQTLs. There were 767 eQTLs that significantly affected leaf expression, pollen expression, and the differences between them, 735 of which affected leaf and pollen expression in the same direction, consistent with the concordance in eQTL effect sizes shown in Figure 2.3b.

## Selective pressure on eQTLs

We examined the distribution of minor allele frequencies (MAF) of eQTLs to test if selective pressure differs for eQTLs between sex or life stages. An excess of low-frequency variants

indicates purifying selection, whereas an excess of intermediate-frequency variants indicates balancing selection (Tajima 1989). We kept one eQTL per eGene to have a set of independent eQTLs for MAF comparisons and repeated the analyses when keeping the top eQTL and a random eQTL. In all tissues, there were more rare alleles than common alleles (Figure 2.4). The median MAFs between tissues were similar, with the highest median MAF in female leaf and lowest median MAF in male leaf using both methods of eQTL selection (Figure 2.4). The differences in mean MAFs between tissues were marginally significant between male and female leaf in top and random eQTLs (Mann-Whitney $U$-test, $p = 0.041$, $0.049$, respectively) and between female leaf and pollen in top eQTLs (Mann-Whitney $U$-test, $p = 0.049$).
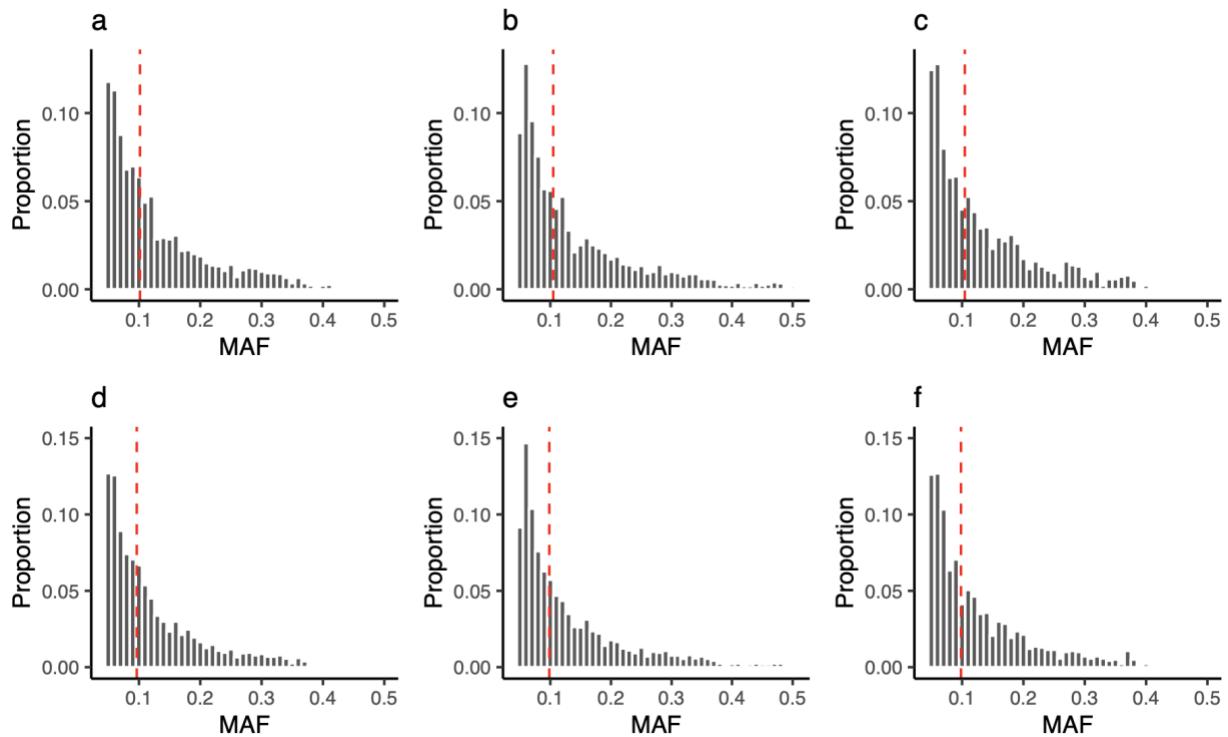


Figure 2.4. Distribution of MAFs of eQTLs in male leaf (a, d), female leaf (b, e), pollen (c, f) in *Rumex hastatulus*. a-c: top eQTL per eGene, d-e: a randomly selected eQTL per eGene. Binwidth = 0.01. Median MAF (dashed lines) in each plot: 0.1017 (a), 0.1048 (b), 0.1045 (c), 0.0965 (d), 0.0984 (e), 0.0982 (f).

To investigate whether the patterns of more rare than common alleles in eQTLs suggested a signature of genome-wide purifying selection, we compared the MAFs of true positive eQTLs to a null distribution in each tissue (Figure 2.5). We generated the null distributions of MAFs using

'false-positive' eQTLs from permuted data by breaking the true assignment of individuals and re-assigning their expression phenotypes. We then used the same procedure to identify false-positive eQTLs from permuted data and true positive eQTLs from observed data (see Methods). In all tissues we found an excess of rare alleles (MAF between 0.05 and 0.1) in the permuted data compared to our observed data, with a small overlap in the proportions of rare alleles in pollen (Figure 2.5), suggesting that eQTLs do not show a genome-wide signature of purifying selection.
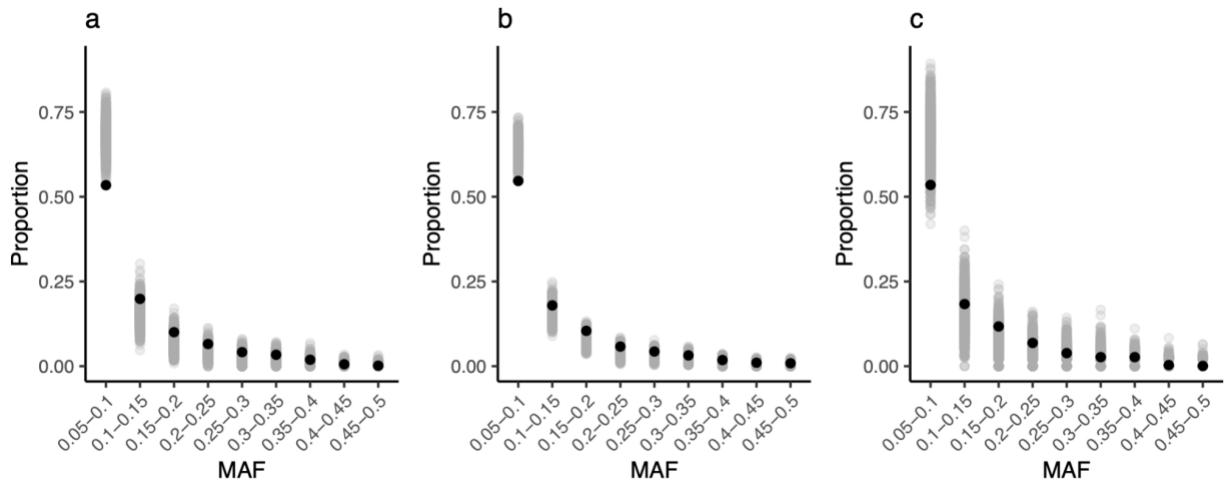


Figure 2.5. Distribution of MAFs of eQTLs compared to a null distribution in male leaf (a), female leaf (b) and pollen (c) in *Rumex hastatulus*. Black dots: true positive eQTLs (observed data), grey dots: false positive eQTLs (permuted data).

We identified 111 eGenes with 714 eQTLs in the PAR, and 1,294 eGenes with 16,112 eQTLs in the X-specific regions in female leaf and compared their selective pressure using MAFs. The MAFs of eQTLs in PAR and the X-specific regions showed an more rare than common alleles similar to the autosomes (Supplementary Figure A8). We found eQTLs in the X-specific regions had significantly lower MAFs than in autosomes using both methods of eQTL selection (Mann-Whitney $U$-test, $p < 0.004$), and eQTLs on PAR had significantly higher MAFs than the X-specific regions only in top eQTLs (Mann-Whitney $U$-test, $p = 0.019$). However, the null distribution of MAFs of eQTLs for X and autosomes did not differ significantly (Supplementary Figure A9); there were not enough false-positive eQTLs in the PAR for comparison.

We compared MAFs of eQTLs shared between life stages and those that were life-stage specific. We found shared eQTLs had significantly higher MAFs than specific eQTLs (Mann-Whitney $U$-

test, $p < 10^{-7}$), with median MAFs of 0.143 and 0.093, respectively. Due to a lack of shared false positive eQTLs between life stages, this result was not confirmed by a comparison to a null distribution. We also tested signatures of selection on eGenes that were shared between sexes or life stages using neutral diversity statistics. We predicted that shared eGenes would have higher $\pi$ and Tajima's $D$ than specific eGenes, due to the greater potential for conflict in shared eGenes. We found that shared eGenes between sexes had significantly higher Tajima's $D$ (Kruskal-Wallis test, $p < 10^{-5}$) but not significantly higher $\pi$ (Kruskal-Wallis test, $p = 0.5383$) than male- or female-specific eGenes. There were no significant differences in $\pi$ or Tajima's $D$ among shared and specific eGenes in either life stage (ANOVA test, $p = 0.611$, 0.157, respectively). Additionally, we tested the signature of selection on eGenes whose expression differences between life stages were affected by eQTLs and found these eGenes had significantly higher $\pi$ and Tajima's $D$ than eGenes not significant for expression differences (*t*-test, $p = 0.00013$, 0.00058, respectively).

## Discussion

In this study we mapped *cis*-eQTLs in leaf and pollen tissues in the dioecious plant *R. hastatulus* to test the scope for conflict between sexes and life stages, based on shared genetic architecture. We investigated the overlap of eQTLs between sexes and life stages and examined their effect sizes and allele frequencies. We discovered a much larger overlap in eGenes and eQTLs between sexes than between life stages. For SNPs tested in both sexes or life stages, there was a stronger correlation of effect sizes between sexes, which further suggests a more shared genetic architecture between sexes than life stages. There were more minor alleles of eQTLs driving than constraining divergent expression between life stages. Among the eQTLs shared between life stages and affecting their expression differences, most minor alleles changed expression in the same direction with different magnitudes, increasing the degree of divergent expression. MAFs of eQTLs in all tissues exhibited similar distributions with a large proportion of rare alleles, but this pattern did not reflect an excess of rare variants when compared to a null distribution. Below we discuss the implications of our results on the extent of a shared genetic basis between sex and life stages and also the use of eQTL mapping in testing evolutionary forces shaping *cis*-regulatory variation.

## *Potential for conflict between sexes and life stages*

With the potential for both sexual and life-stage conflict based on theory (Kidwell et al. 1977; Immler et al. 2012; Peters and Weis 2018), key open questions remain on whether there are more opportunities for conflict between sexes or life stages, and whether sexual and life-stage conflict is still an ongoing process or has already become resolved. We investigated this question using *cis*-regulatory variation which has a strong effect on gene expression. Our results consistently suggested significantly more shared genetic architecture between sexes than life stages, thus more potential for ongoing sexual conflict. In contrast, our results indicate that life-stage conflict is more resolved compared to sexual conflict. This finding is consistent with previous work suggesting considerable scope for ongoing sexual conflict based on intersexual genetic covariances in expression (Innocenti and Morrow 2010; Griffin et al. 2013; Houle and Cheng 2021). The alternating life cycle of diploid and haploid life stages is shared among all land plants, allowing them to adapt to varying environments during their lives. In dioecious plants such as *R. hastatulus*, it is likely that there has been much more time to resolve conflict between life stages than sexes, while in contrast, sexual conflict may still be ongoing. It is also worth mentioning that standing variation in gene regulation might have more sex-differential effects in reproductive tissues such as flowers compared to leaves. Future studies should benefit from a more diverse sampling of both vegetative and reproductive tissues than we used in this study (Puixeu et al. 2023).

With a weak positive correlation of eQTL effect sizes between life stages, there is still potential, albeit weaker, for ongoing life-stage conflict. We found that genes on the Y chromosome with higher expression in pollen compared to leaf also tend to be male-biased in leaves, suggesting the possibility of strong selection in pollen possibly driving up leaf expression in an antagonistic fashion. Future studies with more functional understanding, e.g., on alleles beneficial for pollen competition, combined with candidate genes under balancing selection should help to uncover the genetic basis of life-stage conflict in *R. hastatulus*. Plants are useful systems for comparing opportunities for sexual and life-stage conflict compared to animals; yet studies on selection and potential conflict between sexes or life stages in plants are still very limited (Prasad and Bedhomme 2006; Moore and Pannell 2011; Beaudry et al. 2020). Whether there are differences in the possibilities for sexual and life-stage conflict in plants with different sexual systems

(hermaphroditic vs. dioecious) or different types of life cycles (predominantly diploid vs. haploid) would certainly be worth further investigation.

## *Selective pressure on cis-regulatory variation*

Estimating selective pressures in regulatory variation using allele frequencies has resulted in mixed results supporting contrasting hypotheses about the type of selection on *cis*-eQTLs (Josephs et al. 2015; Brown and Kelly 2021). Our results indicated more rare than common alleles among cis-eQTLs in all tissues. However, in contrast to our prediction and previous findings in *Capsella grandiflora* (Josephs et al. 2015), we found a marginally higher proportion of rare alleles in the null distribution than our observed data, suggesting a genome-wide signature of purifying selection was unlikely. With the *cis*-eQTLs in *Mimulus* floral tissues showing more common than rare alleles (Brown and Kelly 2021), it is unclear whether *cis*-regulatory variation is generally subject to purifying selection based on their allele frequencies. Conserved non-coding regions are often constrained due to their putative regulatory functions (Casillas et al. 2007; Haudry et al. 2013); excessive rare alleles in cis-eQTLs were found in the promoter regions of yeast (Ronald and Akey 2007). Notably, selection on noncoding sequences in plants is rarer than in animals (Williamson et al. 2014), due to low complexity of regulatory regions on a per gene basis (Lockton and Gaut 2005). Investigation of the burden of rare alleles on genotypes with more extreme gene expression (Uzunović et al. 2019) could be a promising next step that might be more powerful than comparing genome-wide allele frequency spectra to a null distribution. Future studies should also benefit from examining the correlation between effect sizes and MAFs while controlling for confirmation bias using allele-specific eQTLs (Josephs et al. 2015) and also examining trans-eQTLs as they are subject to stronger constraint than *cis*-acting variation (Josephs et al. 2020).

In addition to testing selective pressures on eQTLs based on allele frequencies, we examined signatures of selection on eGenes using diversity statistics. We expected eQTLs with concordant effects between sexes or life stages to experience stronger conflict than eQTLs with discordant effects (assuming fitness effects of expression are concordant between sexes or life stages) or eQTLs exclusive to one sex or life stage. We only found that shared eGenes had higher Tajima's *D* than sex-specific eGenes. In contrast to our prediction, eGenes with eQTLs affecting their expression differences between life stages had significantly higher $\pi$ and Tajima's *D* than eGenes

not significant for expression differences; however, this finding might result from an ascertainment bias, as it is more likely to have higher power to detect interactions when diversity is higher. In particular, there was no indication of greater balancing selection on eQTLs with shared genetic architecture between the sexes and life stages. Detecting balancing selection due to conflict using genomic data remains a challenging problem (Flintham et al. 2025; Ruzicka et al. 2025). Additionally, previous work has shown that sexually concordant selection also plays an equally important role in shaping sexual dimorphism in expression (Houle and Cheng 2021). My results from Chapter 3 suggest that concordant selection is likely more prevalent and/or detectable than antagonistic selection in genome-wide studies.

# Methods

## *Plant materials*

We collected open-pollinated seeds from a population in Rosebud, Texas, US (Pickup and Barrett 2013) and planted seedlings in a glasshouse at the Earth Sciences Center, University of Toronto. At the time of flowering, we paired one female and male individual randomly for crossing and collected seeds to generate $F_1$ plants. We planted and retained multiple individuals per family for 78 $F_1$ families to ensure we had one male-female full sibling pair per family for tissue collection. We collected leaf tissue from each male-female full-sibling pair for both RNA and DNA sequencing and collected mature pollen grains from males for RNA sequencing. Growth conditions and leaf tissue collection methods are available in Rifkin et al. (2022). We collected pollen grains from the male $F_1$ plants for RNA isolation based on a modified protocol from Lu (2011). We collected 2-3 mature inflorescences per male individual into a 15 mL falcon tube and added ice-cold 0.5 M mannitol solution till the tube was full. We vortexed the tube vigorously for 1 minute to release the pollen grains, filtered the pollen suspension through a 30 μm nylon mesh, and centrifuged the tube at 450 x g for 5 minutes (4 °C). We repeated the above step with fresh mannitol solution. We transferred the final pollen pellets to a 2 mL centrifuge tube that was flash frozen in liquid nitrogen and stored under -80 °C. We used the standard protocols of Qiagen Plant Mini kit and Sigma Aldrich Spectrum Plant Total RNA Kits for DNA and RNA isolation, respectively. All libraries were prepared and sequenced at the Genome Quebec Innovation Centre in Montréal, Canada. The DNA samples were sequenced at a depth of 10-15X, sample sizes and numbers of reads for DNA and RNA samples were presented in Supplementary Table A1-A2.

## DNA Seq data and variant calling

We used a phased PacBio genome assembly consisting of four autosomes, X-specific, Y-specific and pseudoautosomal regions for our analyses (Sacchi et al., unpublished manuscript). We mapped the DNA reads to the genome assembly and added read groups using bwa-mem2 (Vasimuddin et al. 2019). We sorted the BAM, marked PCR duplicates and indexed the final BAM using SAMTools (Danecek et al. 2021). We called variants in parallel on 2 Mb non-overlapping contigs of each chromosome for all samples using BCFtools mpileup (Danecek et al. 2021), then concatenated the VCFs of contigs for each chromosome. We removed one male sample (24fM) due to low genomic coverage and two samples due to incorrect sex labeling (35aM, 40aF), and only kept samples with both DNA and RNA sequences from the same individual. For males, we only kept samples with both leaf and pollen RNA sequences from the same individual. Our sample sizes for VCF filtering were 74 for female leaf, 73 for male leaf, making a total of 147 leaf tissue samples. We performed SNP filtering as follows for female leaf, male leaf and all leaf samples combined separately using VCFtools (Danecek et al. 2011). We kept SNPs with QUAL > 30, a mean depth between 5 and 20, a genotype quality per sample > 30 otherwise marked as missing, and finally a missingness < 20%. For female samples, we used the same filtering criteria for autosomes, PAR and X-specific regions. The numbers of autosomal SNPs were 64,988,762 in male leaf, 66,828,708 in female leaf, and 66,341,762 in all leaf samples combined. The numbers of SNPs on X-specific regions and PAR in females were 7,745,941 and 2,164,214, respectively. Additionally, we filtered for invariant sites on autosomes with a mean depth between 5 and 20 and a missingness < 20% for all leaf samples combined.

We removed SNPs with a minor allele frequency lower than 5% or with strong Hardy–Weinberg deviations ($p < 10^{-6}$) on autosomes using Plink (Chang et al. 2015) to test for population structure and for the subsequent eQTL mapping. We removed four male samples (7bM, 27eM, 53bM, 5aM) from subsequent analyses for being outliers in principal component analyses in Plink (Chang et al. 2015). Our final sample sizes were 74 for female leaf, 69 for male leaf, making a total of 143 leaf samples. After the filtering using Plink, the numbers of SNPs on autosomes were 3,288,995 in male leaf, 4,509,904 in female leaf, and 3,684,937 in all leaf samples

combined; the numbers of SNPs on X-specific regions and PAR in females were 769,059 and 96,911, respectively.

## RNA Seq data and expression analyses

We mapped the RNA reads to our genome assembly in a two pass mode, sorted the BAM files by coordinates, added read groups using STAR (Dobin et al. 2013), and indexed the output BAM files using SAMTools (Danecek et al. 2021). We generated read counts for each gene using featureCount (Liao et al. 2014). We only kept genes with evidence of expression for our subsequent analyses, i.e., genes with a mean raw read count ≥ 5 in a specific tissue, or in at least one tissue for analyses involving both sexes or life stages. We identified sex-biased genes in leaf tissues and life-stage-biased genes by comparing pollen and male leaf tissues using DESeq2 (Love et al. 2014). The cutoffs for significant differentially expressed (DE) genes were Benjamini-Hochberg adjusted $p < 0.05$ and fold change > 2 unless stated otherwise.

We prepared the expression phenotype data for each tissue separately and for all leaf samples combined for eQTL mapping. We normalized raw read counts by sequencing depth using DESeq2 (Love et al. 2014) and performed quantile normalization using the qqnorm function in R (R Core Team 2022). Additionally, we calculated the difference of normalized read counts in pollen and male leaf (male leaf minus pollen) for each gene and performed quantile normalization of its absolute value using the qqnorm function in R (R Core Team 2022).

## Correlation in gene expression

To characterize the correlation in expression across life stages or sexes, we used Pearson correlations and the final normalized read counts as expression level as described above. For male leaves and pollen, we calculated the correlation of a given gene's expression level in each tissue using 69 male individuals and repeated for 16,411 autosomal genes that had evidence of expression (as described above) to generate a distribution of correlations. These correlations are the same as any phenotypic correlation in behavioral or morphological traits, but for gene expression (Stinchcombe and Kelly 2025). To estimate the across-sex correlation in expression, we took advantage of the sib-structure of our design and the fact that we had male and female sibling pairs. We used the expression level of a gene in a male sample and its female sibling as a pair of observations, repeated for 64 sibling pairs in our design, to estimate the correlation in

expression across siblings for a given gene. We then repeated this for 14,992 genes with evidence of expression in the samples to generate a distribution of across-sex correlation in expression. Note that because we only had a single male and female sample per family, rather than replicates, this is a phenotypic approximation of the across-sex genetic correlation ($r_{mf}$), rather than a genetic correlation. We repeated the analyses for genes regulated by eQTLs (eGenes, see below) in at least one life stage or sex following the same steps.

## eQTL mapping

We performed *cis*-eQTL mapping on autosomes separately for male leaf, female leaf, and pollen tissue. For female leaf, we repeated eQTL mapping on X and PAR using the same approach as the autosomes. For male samples, we performed eQTL mapping using the difference in leaf and pollen expression levels as phenotypes in the same way as in leaf or pollen separately. To control for genetic relatedness, we performed LD pruning in Plink (Chang et al. 2015) and identified genetic principal components (PC) using EIGENSOFT smartpca (Patterson et al. 2006; Price et al. 2006). The significance of each PC was determined by the Tracy-Widom test (Patterson et al. 2006; Sztepanacz and Blows 2017). We included the significant PCs ($p < 0.05$) as covariates in eQTL mapping.

We mapped *cis*-eQTL using tensorQTL (Taylor-Weiner et al. 2019) following its guidelines, and tested pre linkage-pruning SNPs within a 20 kb window to the transcription start site for each gene. We performed adaptive permutations to calculate empirical and beta-approximated empirical *p*-values for each gene (Ongen et al. 2016), and generated nominal *p*-values for all SNP-gene pairs. We calculated the Storey *q*-value and the nominal *p*-value threshold for significant associations based on the beta-approximated empirical *p*-value for each gene using a false discovery rate (FDR) of 0.1 (Benjamini and Hochberg 1995; Storey 2002). We filtered for genes that had *q*-value < 0.1 and at least one eQTL as our eGenes. We filtered for SNPs with a nominal *p*-value smaller than the nominal *p*-value threshold of its eGene as our eQTLs. The minimum *p*-value threshold is 0.0318 for male leaf, 0.0503 for female leaf, and 0.0166 for pollen. We identified conditionally independent eQTLs for eGenes using a forward-backward stepwise regression model in tensorQTL (GTEx Consortium et al. 2017). We defined primary eQTLs as the top ranked eQTL for each eGene, and any additionally ranked eQTLs as secondary

or conditionally independent eQTLs. The effect size of an eQTL was defined as the slope in the linear regression model from tensorQTL.

We kept one eQTL for every eGene, either the most significant (top) eQTL or a randomly selected eQTL when comparing minor allele frequencies (MAF) and when testing the relationship between effect size and the distance to transcription start site of eQTLs. We removed eQTLs and eGenes when an eQTL affected multiple eGenes when comparing MAFs. We performed gene ontology enrichment of eGenes using topGO (Alexa and Rahnenfuhrer 2023), significant GO terms were selected using a $p$-value cutoff of 0.05 based on Fisher's exact tests. To test for the enrichment of expression bias categories among genes whose expression differences between life stages were affected by eQTLs, we grouped DE gene into four bins based on the absolute value of $\log_2$FoldChange ( $\leq 1$, $>1$ and $\leq 2$, $>2$ and $\leq 3$, $>3$) and performed Chi-squared test, where positive Pearson residuals indicate enrichment and negative residuals indicate underrepresentation.

We ran tensorQTL on all 143 leaf samples to test the linear model: Expression ~ Genotype + Sex + Genotype × Sex. We kept the top association for each gene based on the interaction term (-best only), and calculated multiple-testing corrected [39] and Benjamini-Hochberg adjusted $p$-values (pval_adj_bh) following the recommendations from tensorQTL. The cutoff for eQTLs showing significant Genotype × Sex is pval_adj_bh < 0.1.

## Null distributions of eQTL MAFs

To generate a null distribution for MAFs of eQTLs in each tissue, we performed permutations by randomly re-assigning sample ID's in the expression phenotype file. We performed 1000 permutations in both females and males. For female leaves, we performed the same permutations on X and PAR as the autosomes. For male samples, we performed permutations the same way in male leaf and pollen phenotypes and identified the shared and specific eQTLs between male leaf and pollen in each permutation. We identified false-positive eQTLs using the same $p$-value thresholds from the previous analyses on our observed data. We removed eGenes with more than 10 eQTLs, randomly selected an eQTL from each eGene, and removed eQTLs affecting multiple eGenes. We repeated the same procedure of calculating MAF using true positive eQTLs. The proportions of false positive eGenes compared to true positive eGenes were 0.059 in male leaf,

0.105 in female leaf, and 0.0318 in pollen, based on the median number of false positive eGenes in 1000 permutations.

*Diversity statistics*

We used the script codingSiteTypes.py (https://github.com/simonhmartin/genomics_general/blob/master/codingSiteTypes.py, accessed in 2020) to extract 4-fold degenerate site for calculating neutral diversity statistics. We generated 4-fold VCFs including both variant and invariant sites using all leaf samples combined. We used pixy (Korunes and Samuk 2021; Bailey et al. 2025) to calculate nucleotide diversity ($\pi$) and Tajima's $D$ for all autosomal genes. We only kept genes with at least 50 sites in the subsequent analyses of $\pi$ and Tajima's $D$. We removed genes if they were associated with an eQTL that affected multiple eGenes.

## Data availability

Raw sequencing reads are available through the NCBI SRA database under BioProject PRJNA744278. Scripts used in this study are available at https://github.com/SIWLab/Rumex_eQTL.

## Acknowledgement

## References

Alexa A, Rahnenfuhrer J. 2023. topGO: enrichment analysis for Gene Ontology [Internet]. Available from: https://bioconductor.org/packages/topGO/

Arnqvist G, Rowe L. 2005. Sexual conflict. Princeton University Press.

Bailey N, Stevison L, Samuk K. 2025. Correcting for bias in estimates of $\theta_w$ and Tajima's $D$ from missing data in Next-Generation Sequencing. Mol Ecol Resour. e14104.

Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI. 2020. Evolutionary genomics of plant gametophytic selection. Plant Commun. 1(6):100115.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful

approach to multiple testing. J R Stat Soc Ser B Methodol. 57(1):289–300.

Brown KE, Kelly JK. 2021. Genome-wide association mapping of transcriptome variation in *Mimulus guttatus* indicates differing patterns of selection on *cis*- versus *trans*-acting mutations. Genetics. 220(1):iyab189.

Casillas S, Barbadilla A, Bergman CM. 2007. Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. Mol Biol Evol. 24(10):2222–2234.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 4(1):s13742-015-0047–8.

Crowson D, Barrett SCH, Wright SI. 2017. Purifying and positive selection influence patterns of gene loss and gene expression in the evolution of a plant sex chromosome system. Mol Biol Evol. 34(5):1140–1154.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2):giab008.

Davis JR, Fresard L, Knowles DA, Pala M, Bustamante CD, Battle A, Montgomery SB. 2016. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. Am J Hum Genet. 98(1):216–224.

Delph LF, Arntz AM, Scotti-Saintagne C, Scotti I. 2010. The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. Evolution. 64(10):2873–2886.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15–21.

Duffy KJ, Mdlalose ZM, Johnson SD. 2021. Sexual conflict in hermaphroditic flowers of an African aloe. Int J Plant Sci. 182(3):238–243.

Ewing EP. 1977. Selection at the haploid and diploid phases: cyclical variation. Genetics. 87(1):195–207.

Field DL, Pickup M, Barrett SCH. 2013. Ecological context and metapopulation dynamics affect sex-ratio variation among dioecious plant populations. Ann Bot. 111(5):917–923.

Flintham E, Savolainen V, Otto SP, Reuter M, Mullon C. 2025. The maintenance of genetic polymorphism underlying sexually antagonistic traits. Evol Lett. 9(2):259–272..

Grieshop K, Liu MJ, Frost RS, Lindsay MP, Bayoumi M, Brengdahl MI, Molnar RI, Agrawal AF. 2025. Expression divergence in response to sex-biased selection. Mol. Biol. Evol. 42:msaf099.

Griffin RM, Dean R, Grace JL, Rydén P, Friberg U. 2013. The shared genome is a pervasive constraint on the evolution of sex-biased gene expression. Mol Biol Evol. 30(9):2168–2176.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis

Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) Groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. 2017. Genetic effects on gene expression across human tissues. Nature 550:204–213.

Haig D, Wilczek A. 2006. Sexual conflict and the alternation of haploid and diploid generations. Philos Trans R Soc B Biol Sci. 361(1466):335–343.

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet. 45(8):891–898.

Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. 2014. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. Proc Natl Acad Sci. 111(21):7713–7718.

Houle D, Cheng C. 2021. Predicting the evolution of sexual dimorphism in gene expression. Mol Biol Evol. 38(5):1847–1859.

Immler S, Arnqvist G, Otto SP. 2012. Ploidally antagonistic selection maintains stable genetic polymorphism: ploidally antagonistic selection. Evolution. 66(1):55–65.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. Hurst LD, editor. PLoS Biol. 8(3):e1000335.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. Proc Natl Acad Sci. 112(50):15390–15395.

Josephs EB, Lee YW, Wood CW, Schoen DJ, Wright SI, Stinchcombe JR. 2020. The evolutionary forces shaping *cis*- and *trans*-regulation of gene expression within a population of outcrossing plants. Mol Biol Evol. 37(8):2386–2393.

Josephs EB, Stinchcombe JR, Wright SI. 2017. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? New Phytol. 214(1):21–33.

Kidwell JF, Clegg MT, Stewart FM. 1977. Regions of stable equilibria for models of differential selection in the two sexes under random mating. Genetics. 85(1):171–183.

Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol Ecol Resour. 21(4):1359–1368.

Kudaravalli S, Veyrieras J-B, Stranger BE, Dermitzakis ET, Pritchard JK. 2009. Gene expression levels are a target of recent natural selection in the human genome. Mol Biol Evol. 26(3):649–658.

Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution. 34(2):292.

Lankinen Å, Hydbom S, Strandh M. 2017. Sexually antagonistic evolution caused by male–male competition in the pistil. Evolution. 71(10):2359–2369.

Lankinen A, Kiboi S. 2007. Pollen donor identity affects timing of stigma receptivity in *Collinsia heterophylla* (Plantaginaceae): a sexual conflict during pollen competition? Am Nat. 170(6):854–863.

Lee YW, Gould BA, Stinchcombe JR. 2014. Identifying the genes underlying quantitative traits: a rationale for the QTN programme. AoB PLANTS.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30(7):923–930.

Lockton S, Gaut BS. 2005. Plant conserved non-coding sequences and paralogue evolution. Trends Genet. 21(1):60–65.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550.

Lu Y. 2011. RNA isolation from *Arabidopsis* pollen grains. Bio-Protoc. 1(9):e67.

Madjidian JA, Lankinen Å. 2009. Sexual conflict and sexually antagonistic coevolution in an annual plant. PLOS ONE. 4(5):e5477.

Mank JE. 2017. The transcriptional architecture of phenotypic dimorphism. Nat Ecol Evol. 1(1):0006.

Mishra P, Barrera TS, Grieshop K, Agrawal AF. 2024. *Cis*-regulatory variation in relation to sex and sexual dimorphism in *Drosophila melanogaster*. Genome Biol Evol. 16(11):evae234.

Moore JC, Pannell JR. 2011. Sexual selection in plants. Curr Biol CB. 21(5):R176-182.

Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, Parsana P, Kasela S, Balliu B, Viñuela A, et al. 2020. The impact of sex on gene expression across human tissues. Science. 369(6509).

Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. 2016. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 32(10):1479–1485.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLOS Genet. 2(12):e190.

Peters MAE, Weis AE. 2018. Selection for pollen competitive ability in mixed-mating systems. Evolution. 72(11):2513–2536.

Pickup M, Barrett SCH. 2013. The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. Ecol Evol. 3(3):629–639.

Poissant J, Wilson AJ, Coltman DW. 2010. Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. Evolution. 64(1):97–107.

Prasad NG, Bedhomme S. 2006. Sexual conflict in plants. J Genet. 85(3):161–164.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38(8):904–909.

Puixeu G, Macon A, Vicoso B. 2023. Sex-specific estimation of *cis* and *trans* regulation of gene expression in heads and gonads of *Drosophila melanogaster*. G3 GenesGenomesGenetics. 13(8):jkad121.

Puixeu G, Pickup M, Field DL, Barrett SCH. 2019. Variation in sexual dimorphism in a wind-pollinated plant: the influence of geographical context and life-cycle dynamics. New Phytol. 224(3):1108–1120.

R Core Team. 2022. R: a language and environment for statistical computing. Available from: https://www.R-project.org/.

Rifkin JL, Hnatovska S, Yuan M, Sacchi BM, Choudhury BI, Gong Y, Rastas P, Barrett SCH, Wright SI. 2022. Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. Philos Trans R Soc Lond B Biol Sci. 377(1850):20210226.

Ronald J, Akey JM. 2007. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. PLOS ONE. 2(8):e678.

Ruzicka F, Zwoinska MK, Goedert D, Kokko H, Richter X-YL, Moodie IR, Nilén S, Olito C, Svensson EI, Czuppon P, et al. 2025. A century of theories of balancing selection [accessed 2025 May 27]. Available from: https://www.biorxiv.org/content/10.1101/2025.02.12.637871v1

Sacchi BM, Yuan M, Pyne C, Choudhury BI, Gong Y, Barrett SCH, Wright SI. Early regulatory evolution contributes to Y chromosome degeneration in *Rumex hastatulus*. [unpublished manuscript (will be available during revision)]

Sandler G, Beaudry FEG, Barrett SCH, Wright SI. 2018. The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants. Evol Lett. 2(4):368–377.

Schärer L, Janicke T, Ramm SA. 2015. Sexual conflict in hermaphrodites. Cold Spring Harb Perspect Biol. 7(1):a017673.

Signor SA, Nuzhdin SV. 2018. The evolution of gene expression in *cis* and *trans*. Trends Genet TIG. 34(7):532–544.

Stinchcombe JR, Kelly JK. 2025. Measuring natural selection on the transcriptome. New Phytol.

Storey JD. 2002. A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol. 64(3):479–498.

Sztepanacz JL, Blows MW. 2017. Accounting for sampling error in genetic eigenvalues using random matrix theory. Genetics. 206(3):1271–1284.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3):585.

Taylor-Weiner A, Aguet F, Haradhvala NJ, Gosai S, Anand S, Kim J, Ardlie K, Van Allen EM, Getz G. 2019. Scaling computational genomics to millions of individuals with GPUs. Genome Biol. 20(1):228.

Travers SE, Mazer SJ. 2001. Trade-offs between male and female reproduction associated with allozyme variation in phosphoglucoisomerase in an annual plant (*Clarkia unguiculata*: Onagraceae). Evolution. 55(12):2421–2428.

Uzunović J, Josephs EB, Stinchcombe JR, Wright SI. 2019. Transposable elements are important contributors to standing variation in gene expression in *Capsella grandiflora*. Mol Biol Evol. 36(8):1734–1745.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). p. 314–324.

Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI.

2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. PLoS Genet. 10(9):e1004622.

Wyman MJ, Stinchcombe JR, Rowe L. 2013. A multivariate view of the evolution of sexual dimorphism. J Evol Biol. 26(10):2070–2080.

# Chapter 3: Testing for the genomic footprint of conflict between life stages in an angiosperm and moss species

## Abstract

The maintenance of genetic variation by balancing selection is of considerable interest to evolutionary biologists. An important but understudied potential driver of balancing selection is antagonistic pleiotropy between diploid and haploid stages of the plant life cycle. Despite sharing a common genome, sporophytes ($2n$) and gametophytes ($n$) may undergo differential or even opposing selection. Theoretical work suggests antagonistic pleiotropy between life stages can generate balancing selection and maintain genetic variation. Despite the potential for far-reaching consequences of gametophytic selection, empirical tests of its pleiotropic effects (neutral, synergistic, or antagonistic) on sporophytes are generally lacking. Here, we examined the population genomic signals of selection across life stages in the angiosperm *Rumex hastatulus* and the moss *Ceratodon purpureus*. We compared gene expression between life stages and sexes, combined with neutral diversity statistics and the analysis of the distribution of fitness effects. In contrast to what would be predicted under balancing selection due to antagonistic pleiotropy, we found that unbiased genes between life stages were under stronger purifying selection, likely explained by a predominance of synergistic pleiotropy between life stages and strong purifying selection on broadly expressed genes. In addition, we found that 30% of candidate genes under balancing selection in *R. hastatulus* were located within inversion polymorphisms. Our findings provide novel insights into the genome-wide characteristics and consequences of plant gametophytic selection.

# Introduction

Understanding how genetic variation is maintained for traits under selection remains a key open question in evolutionary biology. Balancing selection maintains genetic variation in a population when different alleles are favored in different contexts, e.g., antagonistic selection between sexes or life history stages, hereafter life stages (Kidwell et al. 1977; Immler et al. 2012). Despite persistent theoretical and empirical interest, we still lack empirical tests of the genome-wide importance of balancing selection. The increasing availability of genomic data helps us to understand how different forms of balancing selection shape genetic variation, and also to identify candidate regions under balancing selection (Bitarello et al. 2023; Ruzicka et al. 2025). Here, we test for genome-wide signals of balancing selection generated by antagonistic pleiotropy between life stages using transcriptomic and population genomic data of the angiosperm *Rumex hastatulus* and the moss *Ceratodon purpureus*.

The alternation of life cycles between diploid and haploid phases in sexually reproducing eukaryotes allows for natural selection to occur during both phases (Mable and Otto 1998). In land plants, the haploid phase (gametophyte) and diploid phase (sporophyte) have varying degrees of complexity in the different major clades of land plants. For example, angiosperms generally have the most elaborated sporophytes and the most reduced gametophytes; their gametophytes (i.e., ovules and pollen) are smaller, ephemeral, have fewer cells, and are dependent on sporophytes. Contrastingly, the haploid gametophytes in bryophytes (mosses, liverworts, and hornworts) are multicellular and often perennial. Differential or even opposing selection can result from the differences in genetic, cellular and organismal levels between these two life stages, creating genomic conflicts (Qiu et al. 2012; reviewed in Immler 2019). How selection on gametophytes (gametophytic selection) affects sporophytes is a key unresolved question in plant evolutionary genetics.

The evolutionary consequences of gametophytic selection have been examined both theoretically and empirically (Haldane 1932; Otto et al. 2015; Immler and Otto 2018; summarized in Immler 2019, Beaudry et al. 2020). The haploid nature of the gametophytic phase is predicted to increase the efficacy of selection due to the lack of dominance (i.e., the masking hypothesis), allowing for more efficient removal of recessive deleterious mutations (i.e., purging) and fixation of recessive beneficial mutations (Crow and Kimura 1965; Kondrashov and Crow 1991; Gerstein and Otto

2009). Consistent with this prediction, some studies have found stronger purifying and positive selection in gametophyte-specific than sporophyte-specific genes (Arunkumar et al. 2013; Cervantes et al. 2023; but see Harrison et al. 2019). Stronger selection on gametophyte-expressed genes can also be driven by gametophytic competition (a form of sexual selection) in angiosperms (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et al. 2022). Additionally, gene expression breadth can be a stronger predictor of selection efficacy than the masking effect in angiosperms and bryophytes (Szövényi et al. 2013).

The common genome shared between life stages provides an opportunity to test how gametophytic selection affects the evolution of the sporophyte stage due to pleiotropy between life stages. Plants share a substantial overlap in gene expression between life stages based on both microarray and mRNA sequencing studies (summarized in Beaudry et al. 2020). In ferns and bryophytes where the gametophytes are photosynthetic and exhibit indeterminate growth, gene expression overlap may exceed that in angiosperms, e.g., estimates are ~60% in the angiosperm *Arabidopsis thaliana* (Walbot and Evans 2003; Borg et al. 2009), 97.7% in the fern *Polypodium amorphum* (Sigel et al. 2018), and 85% in the bryophyte *Physcomitrium patens* (Ortiz-Ramírez et al. 2016). Previous studies on gametophytic selection have compared selective forces in gametophyte -*vs.*- sporophyte-specific genes to test for the effects of haploid expression on the distribution of selection coefficients (Arunkumar et al. 2013; Gossmann et al. 2014; Harrison et al. 2019; Gutiérrez-Valencia et al. 2021; Cervantes et al. 2023). However, the patterns of genetic diversity and selection pressures on genes with overlapping expression, which are more likely to be pleiotropic due to the greater potential for competing selection pressures across life stages (Immler et al. 2012; Peters and Weis 2018), have not yet been specifically examined.

In theory, antagonistic pleiotropy between life stages, i.e., intralocus conflict when alleles of a gene have opposite fitness effects between life stages, can cause balancing selection and maintain genetic variation under some conditions (Immler et al. 2012; Peters and Weis 2018). Despite the theoretical support, empirical evidence of antagonistic pleiotropy in angiosperms is limited to a few studies suggesting sexual conflict across life stages, for example, between male gametophyte and female sporophyte during pollen competition (e.g., Travers and Mazer 2001; Lankinen and Kiboi 2007), and between female gametophyte and male sporophyte in a female meiotic drive system (Fishman and Saunders 2008). In bryophytes, where sporophytes are

dependent only on female gametophytes, parent-offspring conflict may occur (Haig and Wilczek 2006; Johnson and Shaw 2016; Shortlidge et al. 2021), but the mechanism and genetic architecture underlying any such conflict remain unclear. Testing the genome-wide prevalence of antagonistic pleiotropy between life stages and its potential to maintain genetic variation is therefore key to understanding the genetic architecture of conflict between plant life stages.

Given the potential for antagonistic pleiotropy between the life stages for genes with overlapping expression, we predicted that if these effects are widespread, we should see population genetic signals of balancing selection more commonly in genes with overlapping expression (Immler et al. 2012; Peters and Weis 2018). An analogous finding is that genes with intermediate rather than extreme expression bias between sexes show patterns of higher genetic diversity consistent with balancing selection, as expected if extreme expression bias reflects resolved rather than ongoing conflict (e.g., Cheng and Kirkpatrick 2016; Wright et al. 2018; Sayadi et al. 2019; summarized in Ruzika et al. 2020). Similarly, we expected greater genetic diversity, indicated by higher nucleotide diversity ($\pi$) and Tajima's $D$, in unbiased or weakly-biased genes between life stages compared to extremely-biased genes or genes exclusive to one life stage. In addition to the more traditionally used diversity statistics, recently developed model-based approaches offer another way to detect balancing selection (Bitarello et al. 2023) and assess enrichment of biased genes between life stages.

We examined gene expression and genetic diversity patterns in the gametophytic and sporophytic life stages of two plant species: the angiosperm *Rumex hastatulus* (Polygonaceae) and the bryophyte *Ceratodon purpureus* (Ditrichaceae). The annual *R. hastatulus* is dioecious, obligately outcrossing, and wind-pollinated with an XY sex chromosome system (Smith 1964; Rifkin et al. 2021). The uniovulate flowers and the large quantity of pollen commonly produced in wind-pollinated angiosperms suggests high potential for gametophytic selection during pollen competition (Friedman and Barrett 2009; Field et al. 2012; Immler 2019). In the dioicous moss *C. purpureus*, the female and male gametophytes possess the U or V sex chromosome, respectively (Carey et al. 2021a), they produce gametes that fuse to form a diploid zygote (i.e., sporophyte, unsexed); the sporophyte matures and ultimately undergoes meiosis, but its entire development occurs on the female gametophyte plant, providing multiple opportunities for antagonistic pleiotropy (Carey et al. 2021b; McDaniel 2023). The contrasts in the length and

complexity of the gametophytic stage in these two species allow us to evaluate the evolutionary consequences of biphasic plant life cycles in a broader context.

We first compared gene expression in gametophytic and sporophytic tissues of *R. hastatulus* and *C. purpureus,* finding a larger overlap in gene expression and a smaller proportion of biased genes between life stages in *C. purpureus* than *R. hastatulus*. We then examined the population genetic signals of balancing selection due to antagonistic pleiotropy between life stages by examining nucleotide diversity and Tajima's *D* across varying degrees of expression bias between life stages. In contrast with our predictions for genome-wide patterns of antagonistic pleiotropy, our results show elevated nucleotide diversity and Tajima's *D* in highly gametophyte-biased genes in both *R. hastatulus* and *C. purpureus*. We conclude that expression levels and breadth were a stronger predictor of selection efficacy than potential competing selective pressures between life stages. Lastly, we identified in both species genomic regions containing hundreds of candidate genes under balancing selection using a composite likelihood ratio test (Cheng and DeGiorgio 2019, 2020, 2022) and examined the biological functions and differential gene expression patterns of these candidate genes.

## Results

### *Gene expression in gametophytes and sporophytes*

We compared gene expression in gametophyte and sporophyte tissues of *R. hastatulus* and *C. purpureus*. In *R. hastatulus*, we sequenced 75 mature pollen samples and 77 male leaf samples representing the gametophyte and sporophyte stages, respectively (described in Chapter 2, Supplementary Table A2). We used only male leaf tissue for assessing sporophytic expression in *R. hastatulus* to avoid confounding effects from potential sexual conflict between leaves of males and females. In *C. purpureus*, we compared expression between 21 sporophyte and 36 gametophyte samples (whole plant, described in Carey et al. 2021a, Supplementary Table B2, see Materials and Methods). In both species, there is a large overlap in gene expression between the life stages (Table 3.1). In *R. hastatulus*, we identified 15,958 gametophyte-expressed genes, consisting of 52% of all annotated genes across the genome, and 89% of genes that were expressed in either life stage. In *C. purpureus*, we identified 24,097 gametophyte-expressed genes, consisting of 69% of all annotated genes and 93% of genes that were expressed in either life stage. The widespread gene expression in gametophytes could suggest the potential for a

large overlap in expression between life stages and an opportunity to test the pleiotropic effects of mutations between gametophytes and sporophytes.

Table 3.1 Summary of samples used and gene expression analyses in *R. hastatulus* and *C. purpureus*. *P*-adj: Benjamini-Hochberg adjusted *p*-value with a false discovery rate of 0.1, FC: fold change in expression, i.e., the ratio of expression level between the two tissues compared. All percentages are based on comparison to the number of expressed genes.

| | Genome size | Sample size of RNA samples | Number of annotated genes | Number of expressed genes | Number of gametophyte-expressed genes | Number of biased genes (padj < 0.1, FC > 2) |
|---|---|---|---|---|---|---|
| *Rumex hastatulus* | 1.7 Gb (2n) | 75 (n), 77 (2n) | 30,641 (2n) | 17,984 | 15,958 (89%) | 5,635 (n, 31%) 8,653 (2n, 48%) |
| *Ceratodon purpureus* | 358 Mb (n) | 36 (n), 21 (2n) | 34,392 (2n) | 25,880 | 24,097 (93%) | 6,446 (n, 25%) 5,202 (2n, 20%) |

We first identified genes that were differentially expressed (DE) between life stages and exclusively expressed in one life stage (Supplementary Table B3). There was a slightly higher percentage of overlapping gene expression and a lower percentage of differentially expressed genes between life stages in *C. purpureus* than *R. hastatulus* (Table 3.1). In *R. hastatulus*, we identified 5,635 gametophyte-biased and 8,653 sporophyte-biased genes; similarly in *C. purpureus*, there were 6,446 gametophyte-biased and 5,202 sporophyte-biased genes (Table 3.1). Among significant DE genes, we filtered for genes in the top 10% quantile of expression bias, allowing us to compare genes with different tissue specificities. Our filtering process resulted in similar numbers of gametophyte- and sporophyte-specific genes: 1,711 gametophyte-specific and 1,711 sporophyte-specific genes in *R. hastatulus*, and 1,886 gametophyte-specific and 1,867 sporophyte-specific genes in *C. purpureus*. Gene ontology (GO) enrichment analyses indicated very different functional enrichment for biased and specific genes between life stages (Supplementary Tables B5, B6), highlighting distinctive expression profiles across life stages in both species. For example, in *R. hastatulus*, gametophyte-specific genes had protein phosphorylation functions (GO:0006468) which are important during pollen-tube growth

(Klodová and Fíla 2021) and sporophyte-specific genes showed photosynthetic functions (GO:0015979).

## *Nucleotide diversity and expression bias between life stages*

Population genomic analyses can be used to provide an indirect test for pleiotropic effects across life stages. We calculated nucleotide diversity at 4-fold ($\pi_s$) and 0-fold ($\pi_n$) sites for each gene and looked at the linear relation between $\pi_s$ or $\pi_n$ and the direction and degree of expression bias between life stages (log2FoldChange) across genes. The linear regression of $\pi_s$ or $\pi_n$ on the direction and degree of expression bias between life stages (log2FoldChange and its absolute value, where positive values indicate gametophyte bias and negative values indicate sporophyte-bias) showed a significant but weak positive relationship, with near-zero slopes in both species (Supplementary Figures B1, B2). We next divided the genes into eight bins based on the direction and degree of expression bias between life stages and calculated the weighted mean of $\pi_s$ and $\pi_n$ for each bin, weighted by the number of sites per gene. In both species, highly gametophyte-biased genes exhibited the highest levels of $\pi_s$, $\pi_n$ and $\pi_n/\pi_s$ ratio (Figure 3.1, Supplementary Table B6). This result contrasts with our expectation that genes with overlapping and intermediate bias in expression should show the highest diversity under a model of widespread antagonistic pleiotropy. In both species, $\pi_s$ and $\pi_n$ increase with the degree of expression bias between life stages in both directions (less so in *R. hastatulus*), but values of $\pi_s$ and $\pi_n$ were greatest for the highly gametophytic-biased genes (Figure 3.1). In *R. hastatulus*, the increase in $\pi_n$ with the degree of expression bias was less symmetrical than in $\pi_s$, with sporophyte-biased genes having lower $\pi_n$ than genes with overlapping expression (Figure 3.1a, 3.1b). Our results were robust to our choice of weighting $\pi_s$ and $\pi_n$ by the number of sites per gene, as when we estimated the mean and standard errors without weighting by the number of sites, the patterns remained the same (Supplementary Figure B3).

There are two possible explanations for the patterns of elevated $\pi_s$ and $\pi_n$ in the most gametophyte-biased genes, a greater prevalence of balancing selection under the assumption of neutrality of synonymous mutations or relaxed purifying selection on this set of genes. If there is relaxed purifying selection on gametophytic-biased genes, this could drive higher $\pi_n$ and higher $\pi_n/\pi_s$. The observed $\pi_n/\pi_s$ ratios showed a similar trend as $\pi_s$ and $\pi_n$, especially in *C. purpureus*,

with highly gametophyte-biased genes having the highest values (Figure 3.1c, 3.1f), suggesting a reduction in the strength of purifying selection on the most gametophyte-biased genes.
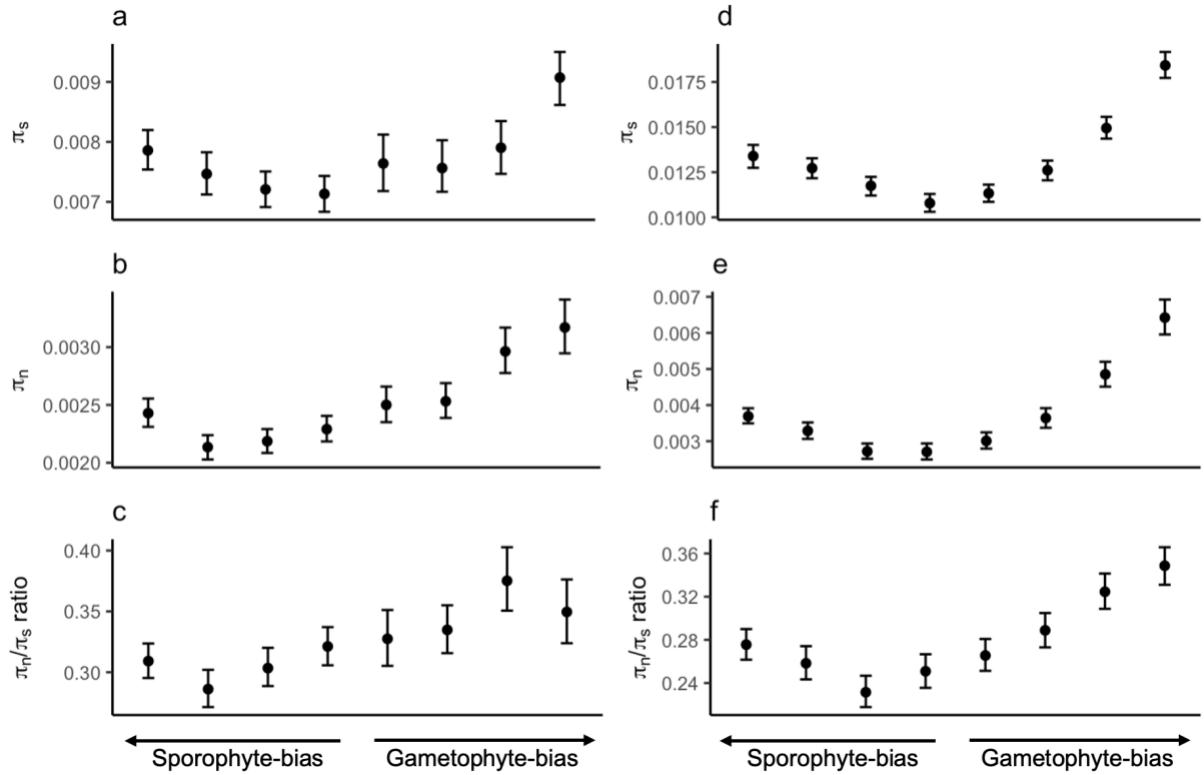


Figure 3.1. Effect of expression bias between life stages on nucleotide diversity in *R. hastatulus* (a-c) and *C. purpureus* (d-f). x-axis: quantiles of log2FoldChange between gametophyte and sporophyte expression. Number of genes in each bin: 1,009-1,847 for *R. hastatulus*; 1,943-2,564 for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 1000 bootstraps of the original dataset. Note that the scales on the y-axis are different between species.

To further distinguish between the two alternative explanations for increased nucleotide diversity in gametophyte biased genes, we examined diversity patterns at genes specifically expressed in one of the two life stages, since these genes should experience the least pleiotropy across stages with reduced expression breadth. We compared $\pi_s$, $\pi_n$ and $\pi_n/\pi_s$ ratio for gametophyte-specific, sporophyte-specific and unbiased genes (Figure 3.2, Supplementary Figure B4, Supplementary Table B7). The results are mostly consistent with Figure 3.1, with gametophyte-specific genes having a higher level of $\pi_s$ and $\pi_n$ than sporophyte-specific and unbiased genes in both species (Figure 3.2a, 3.2b, 3.2d, 3.2e), although the $\pi_n/\pi_s$ ratios were not elevated above unbiased genes in *R. hastatulus* as in *C. purpureus* (Figure 3.2c, 3.2f).
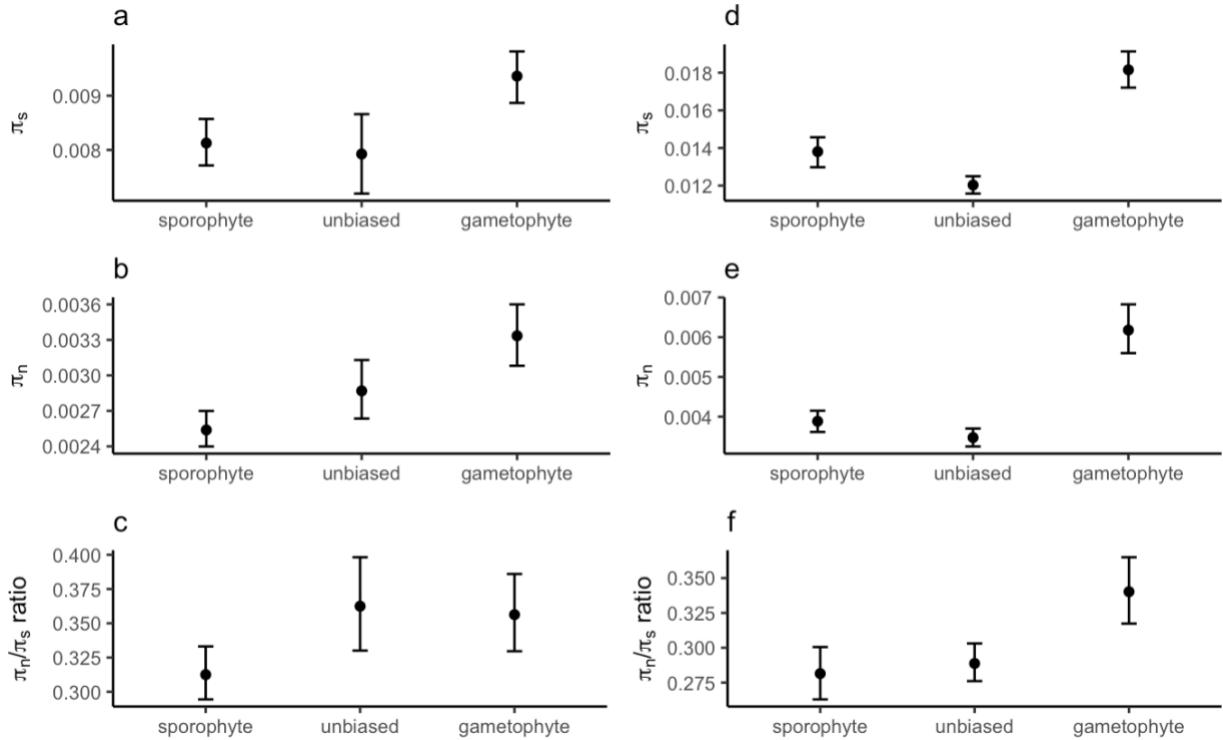
Figure 3.2. Weighted mean nucleotide diversity of gametophyte-specific, sporophyte-specific, unbiased genes in *R. hastatulus* (a-c) and *C. purpureus* (d-f). Number of genes in each category: 1149 (s), 920 (g), 524 (unbiased) for *R. hastatulus*; 1,232 (s), 1,217 (g), 3380 (unbiased) for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 1000 bootstraps of the original dataset. Note that the scales on the y-axis are different between species.

One possible additional factor that could contribute to differences in the strength of purifying selection is if these genes differ in expression level. Gene expression has been shown to be a strong predictor of the strength of purifying selection genome-wide in many species (e.g., Urrutia and Hurst 2003; Slotte et al. 2011; Zhang and Yang 2015). We used this relation to further examine the possible effects of relaxed purifying selection on $\pi_s$, and divided the genes into ten bins of expression level (baseMean across all samples) to examine how expression level affects $\pi_s$, $\pi_n$ and $\pi_n/\pi_s$ in gametophyte- and sporophyte-biased genes. As expression level increases, both $\pi_n$ and $\pi_s$ decrease, with a bigger decrease in $\pi_n$ than $\pi_s$ (Figure 3.3). Gametophyte-biased genes have higher $\pi_s$, $\pi_n$ and $\pi_n/\pi_s$ than sporophyte-biased genes, regardless of expression level in both species, except for the bins with overlapping confidence intervals in $\pi_s$ of *R. hastatulus* (Figure 3.3). Our results suggest that expression level has a strong effect on patterns of diversity. Overall, contrary to our expectations, genes with overlapping expression levels did not show

signals of elevated diversity due to balancing selection. Instead, heterogeneity in diversity likely reflects differences in the strength of purifying selection.
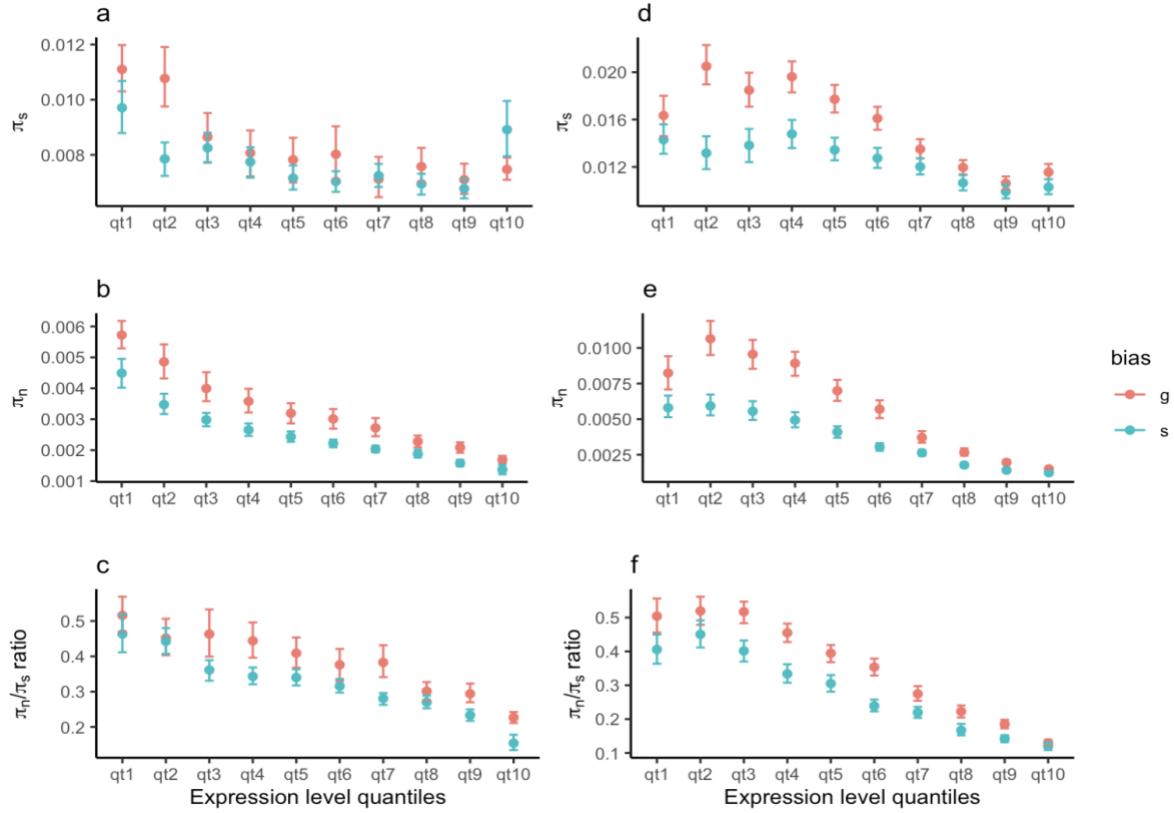


Figure 3.3. Effect of expression level on weighted mean nucleotide diversity in *R. hastatulus* (a-c) and *C. purpureus* (d-f). Number of genes in each bin: 686-1,464 for *R. hastatulus*; 1,034-2,328 for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 500 bootstraps of the original dataset. qt1: lowest expression, qt10: highest expression, g: gametophyte-biased genes, s: sporophyte-biased genes, g: gametophyte-biased genes, s: sporophyte-biased genes. Note that the scales on the y-axis are different between species.

## Tajima's D and expression bias between life stages

Tajima's $D$ is a metric summarizing information of the distribution of allele frequencies in a population (Tajima 1989) and is influenced by both selection and demographic history. Positive values are associated with balancing selection or population contraction and negative values indicate purifying selection or population expansion. We further tested the effect of expression bias between life stages on Tajima's $D$ at synonymous ($D_s$) and nonsynonymous ($D_n$) sites (Figure 3.4). In both species, $D_n$ was smaller than $D_s$, consistent with purifying selection on nonsynonymous sites (Figure 3.4a v.s. 3.4b, 3.4c v.s. 3.4d). In *R. hastatulus*, both $D_n$ and $D_s$

were negative, $D_s$ was higher in gametophyte-biased genes than unbiased and sporophyte-biased genes, as expected if they had a higher proportion of genes under balancing selection (Figure 3.4a). The negative values in $D_n$ and $D_s$ in *R. hastatulus* may possibly be attributed to population expansion, but we currently have no evidence on demographic history. In *C. purpureus*, the patterns of $D_n$ were more similar to $\pi_s$ and $\pi_n$ than $D_s$, both $D_s$ and $D_n$ were higher in highly gametophyte-biased genes than unbiased genes (Figure 3.4c, 3.4d). The higher and more positive Tajima's $D$ values in *C. purpureus* compared to *R. hastatulus* likely resulted from differences in the two species' demographic histories. We also compared $D_s$ and $D_n$ in gametophyte- and sporophyte-biased genes across expression level quantiles. In *R. hastatulus*, $D_s$ and $D_n$ remained roughly constant across expression level quantiles, whereas in *C. purpureus*, $D_s$ and $D_n$ decreased with higher expression (Supplementary Figure B5). Gametophyte- and sporophyte-biased genes had similar levels of $D_s$ and $D_n$ across expression levels in both species. Combined with patterns of $\pi$ (Figures 3.1-3.3) our results are consistent with the hypothesis that gametophyte-biased genes are under relaxed purifying selection.
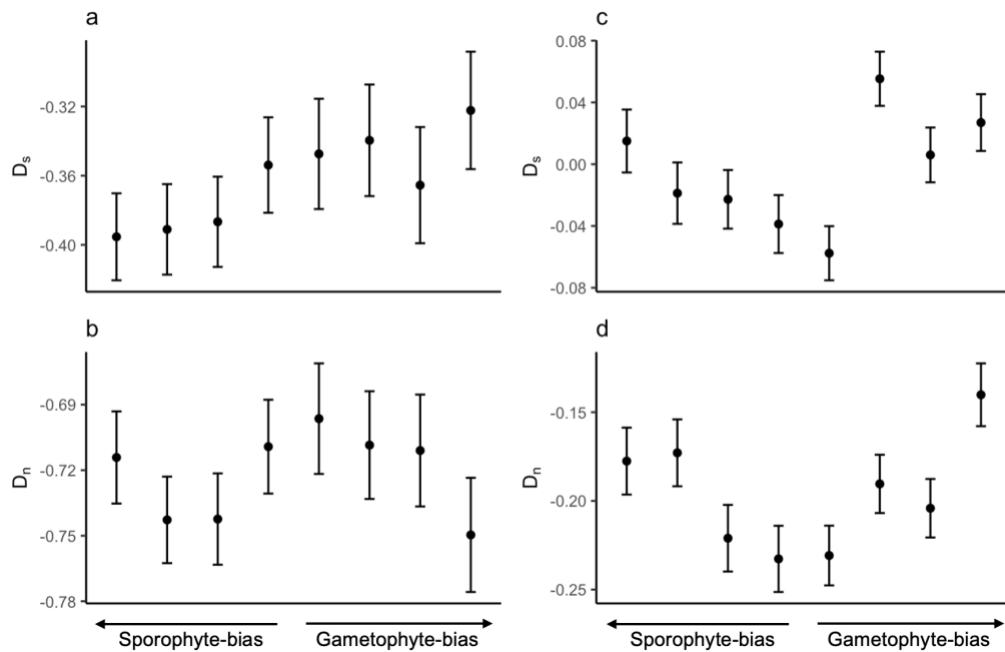


Figure 3.4. Effect of expression bias between life stages on Tajima's $D$ in *R. hastatulus* (a, b) and *C. purpureus* (c, d). x-axis: quantiles of log2FoldChange between gametophyte and sporophyte expression. Number of genes in each bin: 1,009-1,734 ($D_s$), 1,197-1,966 ($D_n$) for *R. hastatulus*; 2,138-2,534 ($D_s$), 2,097-2,579 ($D_n$) for *C. purpureus*. Error bars represent mean ± SEM across genes in each bin. Note that the scales on the y-axis are different between species.

## *Distribution of fitness effects*

Tajima's *D* values reflect a summary of the site frequency spectrum rather than an explicit quantification of the strength of purifying selection on nonsynonymous sites. To address this, we estimated the distribution of fitness effect (DFE) to further understand the strength of purifying selection across life stages while controlling for expression level (Supplementary Methods). We first compared the DFE of gametophyte-specific, sporophyte-specific and unbiased genes (Supplementary Figure B6, Supplementary Table B7), then we binned the gene sets into four expression level quantiles (Supplementary Figure B7, Supplementary Table B9). In both species, higher expression level was associated with lower proportions of effectively neutral mutations ($0 < N_e s < 1$), consistent with an increasing level of constraint due to higher expression level (Supplementary Figure B7).

In *C. purpureus*, unbiased genes had a lower proportion of effectively neutral mutations and a higher proportion of strongly deleterious mutations than gametophyte- or sporophyte-specific genes. This finding is consistent with unbiased genes experiencing higher levels of constraint whereas stage-specific genes are more likely under relaxed purifying selection (Supplementary Figure B6b). Similarly, in the highest expression level quantile (qt4), unbiased genes had a slightly lower proportion of effectively neutral mutations than gametophyte- or sporophyte-specific genes. However, in the other expression level quantiles the same pattern was not evident: unbiased genes had a higher proportion of effectively neutral mutations in the first quantile, and lower proportion of strongly deleterious mutations in the second quantile (Supplementary Figure B7b). In *R. hastatulus*, gametophyte-specific and unbiased genes had similar proportions of effectively neutral and strongly deleterious mutations and sporophyte-specific genes had the lowest proportion of effectively neutral mutations and the highest proportion of strongly deleterious mutations (Supplementary Figure B6a), consistent with Figure 3.2c. Across expression level quantiles, the confidence intervals mostly overlapped between gene sets and mutation categories in *R. hastatulus* (Supplementary Figure B7a).

Based on the haploid purging hypothesis, we predicted that selection efficacy would be stronger in haploid than diploid phases. We therefore expected gametophyte-specific genes to have higher proportions of slightly ($1 < N_e s < 10$) or strongly ($N_e s > 10$) deleterious mutations, or a lower proportion of effectively neutral mutations ($0 < N_e s < 1$). In contrast to previous studies

(Arunkumar et al. 2013; Gossmann et al. 2014; Harrison et al. 2019; Gutiérrez-Valencia et al. 2021; Cervantes et al. 2023), our results in both species did not consistently support the haploid purging hypothesis. In *R. hastatulus*, gametophyte-specific genes had a higher proportion of strongly deleterious mutations than sporophyte-specific genes, while in *C. purpureus* the proportions were similar with overlapping confidence intervals (Supplementary Figure B6). In the low to intermediate expression quantiles (qt1 to qt3), sporophyte-specific genes showed a slightly higher proportion of strongly deleterious mutations ($N_es > 10$) than gametophyte-specific genes in both species (Supplementary Figure B7). In the highest expression quantile (qt4), gametophyte-specific genes had slightly higher proportions of strongly deleterious mutation ($N_es > 10$) than sporophyte-specific genes in both species, but the differences were not significant with overlapping confidence intervals (*R. hastatulus*: 67.19 - 70.21% in n, 65.38 - 68.66% in 2n; *C. purpureus*: 77.09 - 80.36% in n, 74.22 - 78.78 % in 2n, Supplementary Figure B7, Supplementary Table B9).

## Genome-wide scan for balancing selection

These overall patterns of diversity do not show a genome-wide signal of an enrichment of balancing selection on genes with overlapping expression between the life stages; however, it is possible that individual genes may still be subject to balancing selection. We therefore took an alternative approach by conducting a genome-wide scan for signals of balancing selection and explored whether they were enriched for genes that might be subject to antagonistic pleiotropy between life stages. We applied a composite likelihood ratio test to scan the genome for candidate genes under balancing selection based on $B_2$ statistics that account for both increased polymorphism density and an excess of intermediate-frequency alleles (Cheng and DeGiorgio 2019, 2020, 2022). In *R. hastatulus,* a total number of 4,945,561 sites were tested in windows based on genetic positions (Figure 3.5). We found 79,431 sites with strong signals of balancing selection across all chromosomes (CLR > 9.5), located at 449 genes, 405 of which had evidence of expression. The candidate genes were only located in regions with considerable recombination where genetic positions increased rapidly with physical positions, based on our genetic maps (Figure 3.5, Supplementary Figure B11). However, among the genes found in regions with evidence of balancing selection, 138 were located within inversion polymorphisms identified from two phased whole genome assemblies of *R. hastatulus* on chromosomes A2 (63), A3 (74), and A4 (1) (Sacchi et al. 2024; Sacchi et al, in prep), consisting of more than 30% of all

candidate genes; whereas these inversions contained 16.7% of genes across the genome. Because inversion heterozygotes are likely to experience reduced recombination, these candidate genes may be subject to non-independent signals of balanced polymorphism. In *C. purpureus*, the scans were performed in physical position windows as site-specific recombination rates were unavailable (Supplementary Figure B8). We tested 7,743,902 sites and identified 339,375 sites under balancing selection (CLR > 956.2), located at 963 genes, 807 of which had evidence of expression in this study. In *R. hastatulus*, as expected candidate genes under balancing selection had significantly higher $\pi_s$, and significantly lower $\pi_n/\pi_s$ than genes that were not under balancing selection (Welch Two Sample *t*-test, *p*-value = 0.0001511, 0.0003204, respectively). In *C. purpureus*, there was no significant differences in $\pi_s$ or $\pi_n/\pi_s$ between genes under balancing selection and gene that were not (Welch Two Sample *t*-test, *p*-value = 0.6029; Two Sample *t*-test, *p*-value = 0.5956, respectively). We compared the site frequency spectrum (SFS) of all sites tested and the sites under balancing selection (Supplementary Figures B9, B10) and found sites under balancing selection had significantly higher minor allele frequencies than the input data in both species (Mann-Whitney *U* test, *p*-value < 2.2e-16). Note that this is expected, given the use of the site frequency spectrum in the model-based inference. The differences in the shape of SFS in *C. purpureus* is more prominent with a higher proportion of common alleles than in *R. hastatulus* (Supplementary Figure B9, B10).

We performed contingency tests on the number of genes in different expression bias categories (gametophyte-biased, sporophyte-biased, unbiased) for genes under balancing selection and across the whole genome (Supplementary Table B10). In both species, we found no enrichment of any category of expression bias under long-term balancing selection compared to the whole genome (Fisher's exact test, *p* > 0.05). We examined the gene functions for all gametophyte-biased genes in regions under balancing selection (Supplementary Tables B12, B13). We performed GO enrichment after excluding the genes found within inversion polymorphisms in *R. hastatulus*, since the suppression of recombination between inversion types can influence our ability to find functional enrichments of causal loci. The GO enrichment results after excluding inversions were not compelling (Supplementary Table B11). In *R. hastatulus*, two genes under balancing selection had probable disease resistance functions based on orthologs to *Arabidopsis thaliana* (At4g19060, At4g14610). In *C. purpureus*, two genes under balancing selection were

expressed in both life stages with high gametophyte bias and have sexual reproduction functions (log2FoldChange > 3, GO:0019953). These genes were also in the top 10% quantile of $\pi_s$.
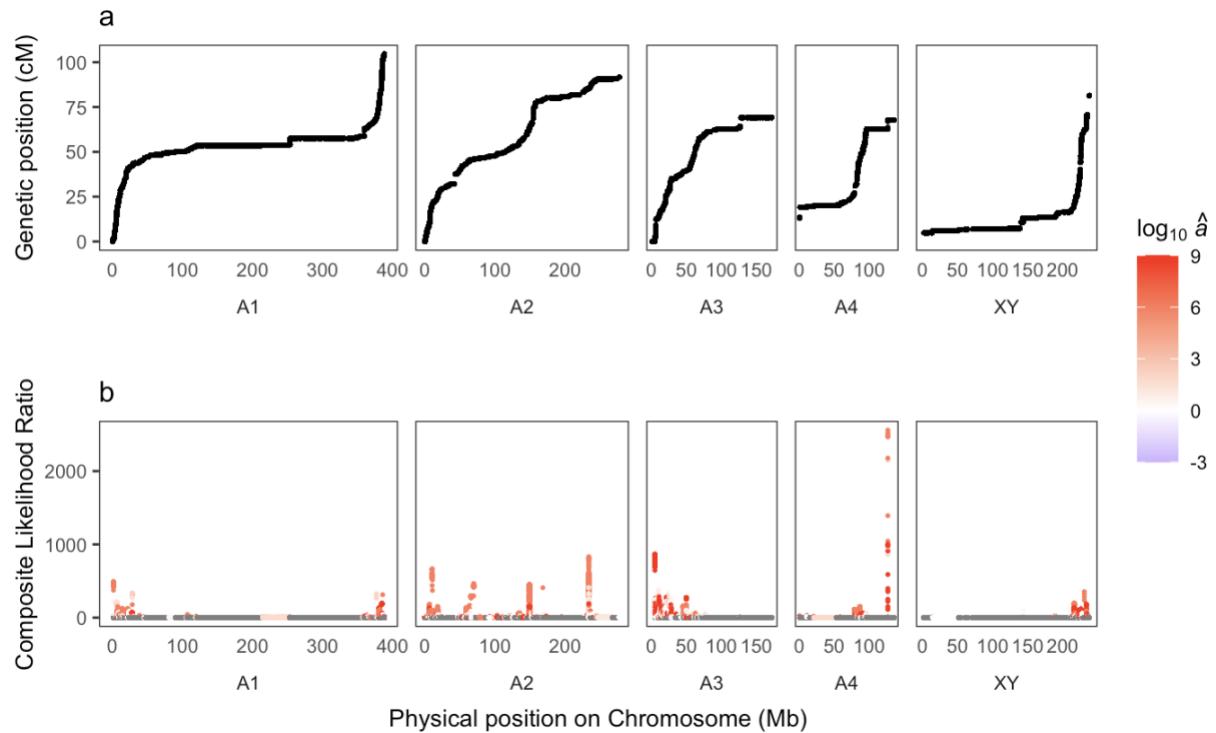


Figure 3.5. Genome-wide scan of balancing selection in *R. hastatulus*. y-axis: genetic positions of sites being tested (a), composite likelihood ratio (b). $\hat{a}$: estimated dispersion parameter, a positive $\log_{10}\hat{a}$ value suggests balancing selection.

## Discussion

We combined population genomic and gene expression analyses to test for balancing selection due to conflict between life stages in *R. hastatulus* and *C. purpureus*. If balancing selection is primarily driven by antagonistic pleiotropy between life stages, we would expect stronger signals of balancing selection in genes expressed in both life stages. Instead, we found elevated $\pi_s$, $\pi_n$ and $D_n$ in highly gametophyte-biased genes in both species. Combined with the observations that $\pi_s$, $\pi_n$ and $\pi_n/\pi_s$ all decreased with increased expression levels, the elevated diversity suggests relaxed purifying selection rather than balancing selection. DFE analyses did not show stronger purifying selection in gametophyte- than sporophyte-specific genes. Lastly, genome-wide scans identified hundreds of candidate genes under balancing selection, but these genes were not enriched for gametophyte-biased expression. Below we discuss how these findings inform our

understanding of plant gametophytic selection and the challenges of alternative approaches for detecting balancing selection.

## *Widespread relaxed purifying selection*

Balancing selection driven by sexual antagonism has been tested using genetic diversity and sex-biased expression (Cheng and Kirkpatrick 2016; Kasimatis et al. 2017; Mank 2017; Wright et al. 2018; Sayadi et al. 2019), showing elevated diversity in weakly sex-biased genes consistent with ongoing conflict. We predicted similar patterns for life stages, however, we found higher $\pi_s$, $\pi_n$ and $D_n$ in highly gametophyte-biased genes (Figure 3.1), which could result from balancing selection or relaxed purifying selection. As highly expressed genes are under strong selective constraint (e.g., Urrutia and Hurst 2003; Slotte et al. 2011; Zhang and Yang 2015), high $\pi_s$ in gametophyte-biased genes with high expression level would suggest balancing selection. Since $\pi_s$ decreased with higher expression level (Figure 3.3), the elevated $\pi_s$ in gametophyte-biased genes suggests relaxed constraint on both nonsynonymous and synonymous sites (Zhang and Qian 2025) is the more plausible explanation rather than widespread intralocus conflict.

Higher $\pi_s$ under relaxed purifying selection could stem from weaker background selection and/or less direct selection on synonymous sites themselves. Background selection reduces neutral nucleotide diversity due to linkage to deleterious mutations, with greater reductions in more constrained regions (Elyashiv et al. 2016). Highly gametophyte-biased genes may be under weaker background selection, elevating $\pi_s$ and $\pi_n$. Though synonymous mutations are generally assumed to be neutral (King and Jukes 1969), studies in fruit flies and yeast showed many synonymous mutations can be deleterious (Lawrie et al. 2013; Shen et al. 2022; Zhang and Qian 2025; but see Kruglyak et al. 2023), likely due to codon usage bias (Plotkin and Kudla 2011; Hunt et al. 2014; Zhou et al. 2016). In *Marchantia polymorpha*, relaxed purifying selection and lower codon usage bias were found in sporophyte-biased genes (Shen et al. 2024). In our case, both synonymous and nonsynonymous sites may be constrained but less so in highly gametophyte-biased genes, resulting in elevated $\pi_s$ and $\pi_n$.

Expression level and breadth may have a stronger effect on selection efficacy than antagonistic pleiotropy across life stages. Our results showed expression level decreased diversity and the proportions of effectively neutral mutations (Figure 3.3, Supplementary Figure B7). Gametophyte- and sporophyte-biased genes likely have more specialized functions and reduced

expression breadth than unbiased genes, particularly pollen-biased genes (Honys and Twell 2003), which could explain their relaxed constraint. However, limited tissue sampling in *R. hastatulus* (pollen and leaf from a specific developmental stage) restricted our assessment of expression breadth and may have affected functional enrichment (Supplementary Table B4). Nonetheless, studies in *Arabidopsis thaliana* have indicated that pollen expression is highly distinct from that of sporophytic tissues, suggesting that pollen-biased genes are likely strongly enriched for gametophytic functions (Rutley and Twell 2015). The lack of female gametophytic tissue (i.e., ovules) in *R. hastatulus* limits our ability to detect potential sexual conflict within the gametophytic stage and between sporophytes and female gametophytes. Additionally, sperm competition can contribute to the maintenance of genetic variation (Clark 2002; Dapper and Wade 2016). Sperm and vegetative cells in pollen might experience different selection pressures, which could explain the patterns of DFE in *R. hastatulus* (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et al. 2022). In contrast, *C. purpureus* expression data included many cell types, the more symmetric expression bias-diversity relationship (Figure 3.1) may reflect smaller differences in expression breadth between gametophytes and sporophytes. Future studies should benefit from a wider and more complete sampling of tissues from different plant species, e.g., including roots in angiosperms like *R. hastatulus*.

Despite different sampling strategies, genome properties, and contrasting predominance of the gametophytic stage, both species showed elevated diversity in highly gametophyte-biased genes. The whole-genome resequencing data of *C. purpureus* came from geographically diverse isolates, while the whole-genome sequencing data of *R. hastatulus* were from a single population (see Material and Methods). Although unlikely to drive a genome-wide pattern, spatially variable selection in *C. purpureus* may maintain different locally adapted alleles especially in the long-lived gametophytes, increasing $\pi$ and Tajima's $D$. The large non-recombining regions in *R. hastatulus* affects gene density and thus patterns of genetic diversity (Rifkin et al. 2022). In contrast, bryophytes like *C. purpureus* show higher recombination rates and more uniform gene density across the genome (Gaut et al. 2007; Bowman et al. 2017; Lang et al. 2018; Healey et al. 2023). The consistent patterns across species support relaxed purifying selection as the most plausible explanation. In contrast to antagonistic pleiotropy, the prevalence of synergistic pleiotropy between life stages may explain the stronger signals of purifying selection on unbiased genes, i.e., deleterious mutations in these genes are selected against in both life stages.

This concordant selection could dominate genome-wide patterns of diversity, and is consistent with breeding studies in which selection on pollen can have synergistic pleiotropic effects on sporophytic fitness (Hormaza and Herrero 1992).

## *Genome-wide scan for balancing selection*

The maintenance of genetic variation through balancing selection is central to evolutionary biology, yet detecting balancing selection genome-wide remains challenging (Ruzicka et al. 2025). Traditional methods rely on the HKA test (Hudson et al. 1987) and summary statistics like Tajima's *D*, which may lack power compared to composite likelihood ratio (CLR) based test (Bitarello et al. 2023). We used the CLR-based *B* statistics, which adjusts window size based on the data and can integrate multiple signatures including an excess of common alleles and increased density of polymorphisms (Cheng and DeGiorgio 2019, 2020, 2022). However, even with improved performance, *B* statistics have limited power when used to detect a single signature of selection and determining appropriate window sizes is challenging without a genetic map. Genome-wide assessments of selection are affected by confounding factors such as demographic history and variation in recombination rates. Limited knowledge of the functional annotation in the study system further complicates the interpretation of candidate sites under selection. Detection power depends on the underlying selection model; for example, Flintham et al. (2025) showed that balancing selection driven by sexual conflict is difficult to detect in genomics, as its polymorphism signature is transient under polygenic selection. Similar challenges are expected when detecting balancing selection due to intralocus conflict between life stages.

Although we did not find a genome-wide signal of intralocus conflict between life stages, we performed functional enrichment of candidate genes under balancing selection to test if they are subject to conflict. We identified 499 and 964 genes under balancing selection in *R. hastatulus* and *C. purpureus*, respectively. Similar approaches should be informative to test whether sexually antagonistic SNPs are under balancing selection (Ruzicka et al. 2019). Our results show over 30% of candidate genes for balancing selection (138 genes) in *R. hastatulus* are located within inversion polymorphisms, consistent with findings from humans (Giner-Delgado et al. 2019). Since recombination is suppressed in inversion heterozygotes, many of these 138 genes may show signals of balancing selection due to linkage disequilibrium rather than being direct

targets of selection. This linkage effect makes it difficult to distinguish focal genes under balancing selection from nearby unselected genes, which potentially affects our functional enrichment results (Supplementary Tables B12, B13). Similarly, we did not observe an enrichment of gametophyte-biased genes among those under balancing selection. However, in *C. purpureus*, we identified two gametophyte-biased genes that show signals of balancing selection; studies of such genes could shed light on the nature of any intergenerational antagonistic pleiotropy.

## *Gene expression profiles in species with biphasic life cycles*

The differentially expressed genes between life stages and their functional enrichment suggest distinct ecological and reproductive roles for gametophytes and sporophytes. In *R. hastatulus*, the greater differentiation in gene expression between life stages (Table 3.1) may reflect the evolutionary reduction of the gametophytic stage in angiosperms, where gametophytes become short-lived, highly specialized, and structurally reduced compared to those in bryophytes (Szövényi et al. 2011). However, differences in RNA sequencing technologies, sample sizes and gene annotations methods between studies limit direct comparisons. Broader comparative analyses across major land plant lineages, such as gymnosperms (Cervantes 2023), using standardized methods would help clarify how life stage complexity and duration affect the extent of overlapping and differential gene expression.

# Materials and Methods

## *Plant materials*

For *R. hastatulus*, we described the details of plant materials and sequencing in Chapter 2; here we provide a brief summary and additional relevant information. We generated seedlings using open-pollinated seeds collected from a population in Rosebud, Texas, US (Pickup and Barrett 2013). After flowering, we paired one female and one male individual randomly and moved them to a mini plastic chamber for crossing. We collected mature seeds from the crosses to generate $F_1$ plants. We collected leaf tissues for DNA and RNA isolation and pollen for RNA isolation using $F_1$ plants, as described in Chapter 2. We used all male leaf and pollen RNA samples from Chapter 2 (Supplementary Table A2) and 20 female leaf DNA samples (Supplementary Table B1). The median read numbers for DNA and RNA sequences are 72,907,022 (range: 64,585,743 - 88,888,320) and 23,412,838 (range: 18,618,763 - 58,151,564),

respectively. Our final sample sizes were 20 female leaf DNA, 77 male leaf RNA, and 75 pollen RNA.

For *C. purpureus*, gametophyte DNA sequences including 18 samples at 6 locations are available from Carey et al. (2021c). All RNA samples were taken from juvenile (protonema) and mature (gametophore) tissues at the gametophyte stage. Gametophyte RNA sequences from 3 male-female sibling pairs at 2 developmental stages (each with 3 biological replicates) are available from Carey et al. (2021a). We generated sporophyte tissues for RNA sequencing using the same methods and at the same time with Carey et al. (2021a). The sporophyte RNA sequences included 26 samples from 6 crosses, the number of replicates per cross is between 2 and 10 (Supplementary Table B2). We removed one sporophyte RNA sample from our analyses because it had low coverage. The median read number for the sporophyte RNA sequences is 52,038,814 (range: 33,030,143 - 64,278,575) (Supplementary Table B2).

## RNA Seq data and expression analysis

For *R. hastatulus*, we used the genome assembly and genome annotation from Rifkin et al. (2022) and assessed the quality of RNA reads using FastQC (Andrews 2010). In *C. purpureus*, we used a hybrid genome consisting of the male R40 genome assembly (autosomes, the V sex chromosome) and the U sex chromosome from the female GG1 genome assembly (Carey et al. 2021a). We trimmed adaptor sequences and filtered out low-quality reads using trimmomatic for all RNA samples (Bolger et al. 2014), and trimmed polyG sequences using fastp for the sporophyte RNA samples (Chen 2023). We used FastQC to assess the quality of RNA reads before and after filtering (Andrews 2010). In *C. purpureus*, we excluded four sporophyte samples for being an outlier in PCA from the majority of sporophyte samples; our final sample size for sporophyte RNA samples is 21.

In both species, we mapped the RNA reads to the genome assembly using STAR in two pass mode (Dobin et al. 2013, v2.7.6a). We sorted the alignment SAM files and added read groups using PicardTools (http://broadinstitute.github.io/picard/). We generated raw read counts for each gene using featureCount (Liao et al. 2014) and then normalized the raw read counts by library depth using DESeq2 (Love et al. 2014). In the expression analysis of *C. purpureus*, we included both male and female gametophyte samples for autosomal genes and only included female or male gametophyte samples for U and V genes, respectively. A gene was identified as

expressed if the mean normalized read count for either life stages was > 5; we only kept expressed genes in the subsequent analyses. We calculated expression ratios between life stages for each gene by dividing the mean normalized read counts at each life stage. The criteria for differentially expressed genes were Benjamini-Hochberg adjusted *p*-value < 0.1, corresponding to a false discovery rate of 0.1 and fold change (FC) >2 (Benjamini and Hochberg 1995), allowing for a sensitive characterization of biological signals and a reasonable control for false positives. A gene was considered tissue-specific if: a) it was in the 10% or 90% quantiles of expression ratio among all significantly differentially expressed genes (adjusted *p*-value < 0.1), and b) it had an absolute value of log2FoldChange > 2. We performed GO enrichment for biased and specific genes using topGO, with expressed genes from each species serving as the gene universe for comparison. (Alexa and Rahnenfuhrer 2023). We divided the genes into four bins based on quantiles of log2FoldChange for gametophyte-biased (log2FoldChange > 0) and sporophyte-biased (log2FoldChange > 0) genes separately. We defined unbiased genes as the middle two bins with lowest absolute value of log2FoldChange as well as an adjusted *p*-value ≥ 0.1. We divided the expressed genes into four or ten quantiles of expression levels and used the same cutoff for later analyses.

## *DNA Seq data and variant calling*

For *R. hastatulus*, we assessed the quality of DNA raw reads using FastQC (Andrews 2010) and mapped the reads to the genome assembly using bwa-mem2 (Vasimuddin et al. 2019). We sorted the BAM files using SAMtools (Danecek et al. 2021), added read groups, and removed PCR duplicates using PicardTools (http://broadinstitute.github.io/picard/). We called SNPs jointly on all samples using freebayes (Garrison and Marth 2012) with invariant sites saved in the output (--report-monomorphic). The variant and invariant sites were filtered separately using VCFtools (Danecek et al. 2011), and concatenated together using BCFtools (Danecek et al. 2021). For both variant and invariant sites we removed sites with a proportion of missing data > 0.8 or with a mean read depth per site <10 or >40. For variant sites, we only kept bi-allelic SNPs with genotype quality > 30.

For *C. purpureus*, we removed adaptor sequences using trimmomatic (LEADING:3 TRAILING:3 SLIDINGWINDOW:10:30 MINLEN:40) (Bolger et al. 2014), and performed quality control using FastQC before and after trimming (Andrews 2010). We masked the V sex

chromosome when mapping female samples and masked the U sex chromosome when mapping male samples using BEDTools maskfasta (Quinlan and Hall 2010). The read mapping and BAM file processing methods are the same as described above in *R. hastatulus*. We called SNPs on all samples jointly using BCFtools mpileup with a ploidy of 1 and the -B option (Danecek et al. 2021). We only kept the VCF for autosomal genes for the subsequent analysis, as the U and V sex chromosomes have low recombination (Carey et al. 2021a, 2021c). We performed VCF filtering using BCFtools (Danecek et al. 2021). We removed sites with a combined read depth < 5 and filtered variant and invariant sites separately. For variant sites, we kept bi-allelic SNPs and removed sites with low quality or mapping score (QUAL<30 && MQ<30). Finally, we removed sites with a proportion of missing data > 0.8 for both invariant and variant sites.

*Diversity statistics*

In both species we identified 0-fold and 4-fold degenerate sites in the genome using the script codingSiteTypes.py (https://github.com/simonhmartin/genomics_general/blob/master/codingSiteTypes.py, accessed in 2020). We intersected the list of 0-fold and 4-fold sites with the filtered VCFs to generate 0-fold and 4-fold VCFs for population genetic analyses. The numbers of sites (variant and invariant) in each VCF are 11,432,804 (0-fold) and 2,891,775 (4-fold) in *R. hastatulus*; 17,394,329 (0-fold) and 4,759,819 (4-fold) in *C. purpureus*. We used pixy to calculate the 0-fold and 4-fold average per site nucleotide diversity for each gene across the genome (Korunes and Samuk 2021, v1.2.6.beta1). Only genes with a minimum of 50 sites for both 4-fold and 0-fold nucleotide diversity were kept in the subsequent analyses.

We used the script popgenWindows.py (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py, accessed in 2020) to calculate Tajima's *D* for each gene based on the variant-only VCFs. We calculated the weighted average nucleotide diversity for each bin based on the number of sites for each gene (see supplementary methods). We bootstrapped the genes in each bin 1000 times to generate 95% confidence intervals. We calculated the weighted average nucleotide diversity (mean and 95% CI) in the same way for gametophyte-specific, sporophyte-specific and unbiased genes. We removed zeros for $\pi_s$ and $\pi_n$ to calculate the $\pi_n/\pi_s$ ratio for each gene.

All DNA samples in *R. hastatulus* are female, this reduces the bias in nucleotide diversity estimates due to the divergence between the X and Y sex chromosomes in males. The DNA samples in *C. purpureus* were generated from 8 males and 8 females from 5 locations, see sample information from (2021a).

*Balancing selection scan*

We used BalLeRMix+ with default settings to perform genomic scans for balancing selection based on $B_2$ statistics, incorporating derived allele frequency, polymorphism and divergence information (Cheng and DeGiorgio 2019, 2020, 2022). We used an outgroup to infer the ancestral state for both species (supplementary methods). A high CLR indicates strong signal for balancing selection. In *R. hastatulus*, the scans were performed on each chromosome using the genetic positions. In *C. purpureus*, the scans were performed in 1 Mb sliding windows based on the physical positions on each chromosome, using a default recombination rate of 1 cM/Mb. The threshold for signature of balancing selection was set as the 5% quantile in the non-zero values of composite likelihood ratio. We calculated the maximum CLR per gene. We identified the genes that had at least one site with strong signal of balancing selection and also with evidence of expression in this study. We performed GO enrichment for candidate balancing selection genes using topGO, with expressed genes as the gene universe for comparison (Alexa and Rahnenfuhrer 2022). For candidate balancing selection genes with gametophyte bias, we used gametophyte-biased genes as the comparison to test the functional enrichment for balancing selection.

## Data availability

## Acknowledgements

# References

Alexa A, Rahnenfuhrer J. 2023. topGO: Enrichment analysis for gene ontology [Internet]. [accessed 2022 May 20]. Available from: https://bioconductor.org/packages/topGO/

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data [Internet]. [accessed 2017 Dec 21]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. Mol Biol Evol. 30(11):2475–2486.

Beaudry FE, Rifkin JL, Barrett SC, Wright SI. 2020. Evolutionary genomics of plant gametophytic selection. Plant Commun.:100115.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 57(1):289–300.

Bitarello BD, Brandt DYC, Meyer D, Andrés AM. 2023. Inferring balancing selection from genome-scale data. Genome Biol Evol. 15(3):evad032.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120.

Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective. J Exp Bot. 60(5):1465–1478.

Bowman JL et al. 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell. 171:287-304.e15.

Carey SB, Jenkins Jerry, Lovell John T., Maumus Florian, Sreedasyam Avinash, Payton Adam C., Shu Shengqiang, Tiley George P., Fernandez-Pozo Noe, Healey Adam, et al. 2021a. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. Sci Adv. 7(27):eabh2488.

Carey SB, Kollar LM, Mcdaniel SF. 2021b. Does degeneration or genetic conflict shape gene content on UV sex chromosomes? Bryophyte Divers Evol. 43(1):133–149.

Carey SB, Peniston JH, Payton AC, Kim M, Lipzen A, Bauer D, Lail K, Daum C, Barry K, Jenkins J, et al. 2021c. Novel insights into joint estimations of demography, mutation

rate, and selection using UV sex chromosomes [Preprint]. bioRxiv.

Cervantes S, Kesälahti R, Kumpula TA, Mattila TM, Helanterä H, Pyhäjärvi T. 2023. Strong purifying selection in haploid tissue–specific genes of scots pine supports the masking theory. Mol Biol Evol. 40(8):msad183.

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2(2):e107.

Cheng C, Kirkpatrick M. 2016. Sex-specific selection and sex-biased gene expression in humans and flies. Nachman MW, editor. PLOS Genet. 12(9):e1006170.

Cheng X, DeGiorgio M. 2019. Detection of shared balancing selection in the absence of trans-species polymorphism. Mol Biol Evol. 36(1):177–199.

Cheng X, DeGiorgio M. 2020. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. Mol Biol Evol. 37(11):3267–3291.

Cheng X, DeGiorgio M. 2022. BalLeRMix+: mixture model approaches for robust joint identification of both positive selection and long-term balancing selection. Bioinformatics. 38(3):861–863.

Clark AG. 2002. Sperm competition and the maintenance of polymorphism. Heredity. 88(2):148–153.

Crow JF, Kimura M. 1965. Evolution in sexual and asexual populations. Am Nat. 99(909):439–450.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2):giab008.

Dapper AL, Wade MJ. 2016. The evolution of sperm competition genes: the effect of mating system on levels of genetic variation within and between species. Evol Int J Org Evol. 70(2):502–511.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15–21.

Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, Andolfatto P, Coop G, Sella G. 2016. A genomic map of the effects of linked selection in *Drosophila*. PLOS Genet. 12(8):e1006130.

Field DL, Pickup M, Barrett SCH. 2012. The influence of pollination intensity on fertilization success, progeny sex ratio, and fitness in a wind-pollinated, dioecious plant. Int J Plant Sci. 173(2):184–191.

Fishman L, Saunders A. 2008. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. Science. 322(5907):1559–1562.

Flintham E, Savolainen V, Otto SP, Reuter M, Mullon C. 2025. The maintenance of genetic polymorphism underlying sexually antagonistic traits. Evol Lett. 9(2):259–272.

Friedman J, Barrett SC. 2009. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. Ann. Bot. 103:1515–1527.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing [Preprint]. arXiv:1207.3907 [q-bio.GN].

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. 8:77–84.

Gerstein AC, Otto SP. 2009. Ploidy and the causes of genomic evolution. J Hered. 100(5):571–581.

Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, Pantano L, Bitarello BD, Izquierdo D, Noguera I, et al. 2019. Evolutionary and functional impact of common polymorphic inversions in the human genome. Nat Commun. 10(1):4222.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. Mol Biol Evol. 31(3):574–583.

Gutiérrez-Valencia J, Fracassetti M, Horvath R, Laenen B, Désamore A, Drouzas AD, Friberg M, Kolář F, Slotte T. 2022. Genomic signatures of sexual selection on pollen-expressed genes in *Arabis alpina*. de Meaux J, editor. Mol Biol Evol. 39(1):msab349.

Haig D, Wilczek A. 2006. Sexual conflict and the alternation of haploid and diploid generations. Philos Trans R Soc B Biol Sci. 361(1466):335–343.

Haldane JBS. 1932. The causes of evolution. Princeton (NJ): Princeton University Press.

Harrison MC, Mallon EB, Twell D, Hammond RL. 2019. Deleterious mutation accumulation in *Arabidopsis thaliana* pollen genes: a role for a recent relaxation of selection. Genome Biol Evol. 11(7):1939–1951.

Healey AL et al. 2023. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. Nat. Plants. 9:238–254.

Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol. 132(2):640–652.

Hormaza JI, Herrero M. 1992. Pollen selection. Theoret Appl Genetics. 83(6–7):663–672.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116(1):153–159.

Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. 2014. Exposing synonymous mutations. Trends Genet. 30(7):308–321.

Immler S. 2019. Haploid selection in "diploid" organisms. Annu. Rev. Ecol. Evol. Syst. 50:219–236.

Immler S, Arnqvist G, Otto SP. 2012. Ploidally antagonistic selection maintains stable genetic polymorphism: ploidally antagonistic selection. Evolution. 66(1):55–65.

Immler S, Otto SP. 2018. The evolutionary consequences of selection at the haploid gametic stage. Am. Nat. 192:241–249.

Johnson MG, Shaw AJ. 2016. The effects of quantitative fecundity in the haploid stage on reproductive success and diploid fitness in the aquatic peat moss *Sphagnum*

*macrophyllum*. Heredity. 116(6):523–530.

Kasimatis KR, Nelson TC, Phillips PC. 2017. Genomic signatures of sexual conflict. J Hered. 108(7):780–790.

Kidwell JF, Cleggi MT, Stewart FM. 1977. Regions of stable equilibria for models of differential selection in the two sexes under random mating. Genetics. 85(1):171–183.

King JL, Jukes TH. 1969. Non-darwinian evolution: most evolutionary change in proteins may be due to neutral mutations and genetic drift. Science. 164:788–798.

Klodová B, Fíla J. 2021. A decade of pollen phosphoproteomics. Int J Mol Sci. 22(22):12212.

Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? Nature. 351(6324):314–315.

Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol Ecol Resour. 21(4):1359–1368.

Kruglyak L, Beyer A, Bloom JS, Grossbach J, Lieberman TD, Mancuso CP, Rich MS, Sherlock G, Kaplan CD. 2023. Insufficient evidence for non-neutrality of synonymous mutations. Nature. 616(7957):E8–E9.

Lang D et al. 2018. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. Plant J. Cell Mol. Biol. 93:515–533.

Lankinen A, Kiboi S. 2007. Pollen donor identity affects timing of stigma receptivity in *Collinsia heterophylla* (Plantaginaceae): a sexual conflict during pollen competition? Am Nat. 170(6):854–863.

Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at synonymous sites in *D. melanogaster*. PLOS Genet. 9(5):e1003527.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma Oxf Engl. 30(7):923–930.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550.

Mable BK, Otto SP. 1998. The evolution of life cycles with haploid and diploid phases. BioEssays. 20:453–462.

Mank JE. 2017. Population genetics of sexual conflict in the genomic era. Nat Rev Genet. 18(12):721–730.

McDaniel SF. 2023. Divergent outcomes of genetic conflict on the UV sex chromosomes of *Marchantia polymorpha* and *Ceratodon purpureus*. Curr Opin Genet Dev. 83:102129.

Ortiz-Ramírez C, Hernandez-Coronado M, Thamm A, Catarino B, Wang M, Dolan L, Feijó JA, Becker JD. 2016. A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. Mol Plant. 9(2):205–220.

Otto SP, Scott MF, Immler S. 2015. Evolution of haploid selection in predominantly diploid organisms. Proc. Natl. Acad. Sci. 112:15952–15957.

Peters MAE, Weis AE. 2018. Selection for pollen competitive ability in mixed-mating systems. Evolution. 72(11):2513–2536..

Pickup M, Barrett SCH. 2013. The influence of demography and local mating environment on

sex ratios in a wind-pollinated dioecious plant. Ecol Evol. 3(3):629–639.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 12(1):32–42.

Qiu Y, Taylor AB, McManus HA. 2012. Evolution of the life cycle in land plants. J Syst Evol. 50(3):171–194.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842.

Rifkin JL, Beaudry FEG, Humphries Z, Choudhury BI, Barrett SCH, Wright SI. 2021. Widespread recombination suppression facilitates plant sex chromosome evolution. Mol Biol Evol. 38(3):1018–1030.

Rifkin JL, Hnatovska S, Yuan M, Sacchi BM, Choudhury BI, Gong Y, Rastas P, Barrett SCH, Wright SI. 2022. Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. Philos Trans R Soc Lond B Biol Sci. 377(1850):20210226.

Rutley N, Twell D. 2015. A decade of pollen transcriptomics. Plant Reprod. 28(2):73–89.

Ruzicka F et al. 2020. The search for sexually antagonistic genes: practical insights from studies of local adaptation and statistical genomics. Evol. Lett. 4:398–415.

Sacchi B, Humphries Z, Kružlicová J, Bodláková M, Pyne C, Choudhury BI, Gong Y, Bačovský V, Hobza R, Barrett SCH, et al. 2024. Phased assembly of neo-sex chromosomes reveals extensive y degeneration and rapid genome evolution in *Rumex hastatulus*. Mol Biol Evol. 41(4):msae074.

Sayadi A, Martinez Barrio A, Immonen E, Dainat J, Berger D, Tellgren-Roth C, Nystedt B, Arnqvist G. 2019. The genomic footprint of sexual conflict. Nat Ecol Evol. 3(12):1725–1730.

Shen T, Gadient P, Goodrich J, Becher H. 2024. Selection efficacy differs between lifestages in the haploid-diploid *Marchantia polymorpha* subsp. ruderalis. :2024.09.06.611587.

Shen X, Song S, Li C, Zhang J. 2022. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature. 606(7915):725–731.

Shortlidge EE, Carey SB, Payton AC, McDaniel SF, Rosenstiel TN, Eppley SM. 2021. Microarthropod contributions to fitness variation in the common moss *Ceratodon purpureus*. Proc R Soc B Biol Sci. 288(1947):20210119.

Sigel EM, Schuettpelz E, Pryer KM, Der JP. 2018. Overlapping patterns of gene expression between gametophyte and sporophyte phases in the fern *Polypodium amorphum* (Polypodiales). Front Plant Sci. 9:1450.

Slotte T, Bataillon T, Hansen TT, St. Onge K, Wright SI, Schierup MH. 2011. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. Genome Biol Evol. 3:1210–1219.

Smith BW. 1964. The evolving karyotype of *Rumex hastatulus*. Evolution. 93–104.

Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ. 2011. Generation-biased gene expression in a bryophyte model system. Mol Biol Evol. 28(1):803–812.

Szövényi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. 2013. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. Mol Biol Evol. 30(8):1929–1939.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585.

Travers SE, Mazer SJ. 2001. Trade-offs between male and female reproduction associated with allozyme variation in phosphoglucoisomerase in an annual plant (*Clarkia unguiculata*: Onagraceae). Evolution. 55(12):2421–2428.

Urrutia AO, Hurst LD. 2003. The signature of selection mediated by expression on human genes. Genome Res. 13(10):2260–2264.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). p. 314–324.

Walbot V, Evans MMS. 2003. Unique features of the plant life cycle and their consequences. Nat Rev Genet. 4(5):369–379.

Wright AE, Fumagalli M, Cooney CR, Bloch NI, Vieira FG, Buechel SD, Kolm N, Mank JE. 2018. Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. Evol Lett. 2(2):52–61.

Zhang J, Qian W. 2025. Functional synonymous mutations and their evolutionary consequences. Nat Rev Genet.:1–16.

Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. Nat Rev Genet. 16(7):409–420.

Zhou Z, Dang Y, Zhou M, Li L, Yu C, Fu J, Chen S, Liu Y. 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc Natl Acad Sci. 113(41):E6117–E6125.

# Chapter 4: Testing for pollen competition and pollen drive in *Rumex hastatulus*

*This chapter was written in collaboration with John R. Stinchcombe and Stephen I. Wright. All authors conceptualized the study, I performed the experiment and data analysis.*

## Abstract

The potential conflict between plant life stages may have important evolutionary consequences. Despite sharing a common genome, the sporophytic (2n) and gametophytic (n) life stages may undergo differential or even opposing selection. In flowering plants, genes beneficial for pollen competition may reduce the fitness of the diploid offspring, creating a conflict that may generate balancing selection. It is important to identify and characterize genetic variation contributing to pollen competition success and test whether they show genetic trade-offs between life stages. We combined experimental crosses and pooled sequencing of pollen grains and early seeds from a single population of the dioecious and wind-pollinated annual *Rumex hastatulus* to test for non-Mendelian transmission due to male meiotic drive and pollen competition. We found that, unbeknownst to us, the pools of early seeds we collected were likely to be primarily unfertilized ovules, preventing an assessment of the genomic extent of pollen competition from this experiment. Allele frequencies differ between leaf and pollen in male plants, but we did not find clear signals of allele frequency distortion caused by male meiotic drive. We found signals of a weak X-bias in pollen grains based on sequencing coverage, suggesting that sex ratio bias may in part arise from X-drive during pollen formation. Our study highlights the challenges and key considerations in using experimental crosses and pooled sequencing to detect genetic variation underlying pollen competition and drive within a single large population, overcoming these technical challenges will be important in future studies.

# Introduction

Pollen competition, a mechanism for sexual selection in plants, has great evolutionary consequences (Mulcahy and Mulcahy 1987; Williams and Mazer 2016; Immler 2019). After pollination, pollen grains germinate on the surface of stigmas and grow pollen tubes that compete to fertilize ovules. Early studies demonstrated evidence for standing variation in male siring success in natural populations (Walsh and Charlesworth 1992). Selection in the gametophytic stage (i.e., pollen and ovules) is predicted to be more efficient than the diploid stage given the absence of dominance (Crow and Kimura 1965; Kondrashov and Crow 1991). Despite the biological significance of gametophytic selection during pollen competition (Hormaza and Herrero 1992; Beaudry et al. 2020), the genomic locations and extent of standing variation affecting pollen competition remain unknown. Additionally, selfish genetic elements in pollen creating segregation distortion can also impact the success of 'drive' in pollen grains (Lindholm et al. 2016). Here, we combined experimental crosses and pooled sequencing to investigate the genetics of pollen competition and pollen drive within a single large population.

Pollen competition is an important component of male siring success. Heritable variation in traits associated with male siring success such as pollen tube growth rate have been shown in different plant species (Walsh and Charlesworth 1992; Beaudry et al. 2020). Past studies assigned paternity of seeds to estimate male contribution to the offspring, often through a small number of microsatellite markers (reviewed in Bernasconi 2003). For example, paternity analysis in *Brassica rapa* found evidence for selection on flowering time through male fitness, likely due to a correlation between age at flowering and pollen quality (Austen and Weis 2016a, 2016b). Recent availability of genomic data makes it possible to use thousands of genetic markers for assigning paternity. In *Mimulus guttatus*, hundreds of SNPs were found to undergo selection from male siring success (Monnahan et al. 2021). These studies strongly supported the potential for gametophytic selection during pollen competition.

Selection during pollen competition has been shown to affect sex ratio bias and sex chromosome evolution in dioecious plants (Delph 2019). Female-biased sex ratios can result from greater success of X- compared with Y-bearing pollen grains in species with X/Y sex determination (Correns 1928; Lloyd 1974), potentially due to Y chromosome degeneration (Hough et al. 2014). In several dioecious *Rumex* species with heteromorphic sex chromosomes, a higher pollen

density is associated with more female-biased sex ratios in the seeds, likely due to stronger pollen competition. Additionally, gametophytic selection can purge mutation load from the deleterious effects of Y chromosome degeneration, and thus slow down Y degeneration (Chibalina and Filatov 2011). Gametophytic selection is also expected to lead to more pollen-beneficial alleles on the Y (Scott and Otto 2017), potentially contributing to the suppression of recombination. Consistent with this prediction, pollen-biased genes were enriched on the Y chromosome compared to autosomes in two *Rumex* species (Sandler et al. 2018), but see Chapter 2.

Meiotic drive occurs when a selfish genetic element increases its segregation ratio more than Mendelian expectation during meiosis (Lindholm et al. 2016). In addition to pollen competition, meiotic drive during gamete production can create strong selection and affect allele frequencies in the seeds. Male meiotic drive is an important force shaping the genetics and evolution of spermatogenesis (Presgraves 2009). In *Silene latifolia*, sex-linked meiotic drivers have been suggested to cause female-biased sex ratios (Taylor 1996, 1999; Taylor and Ingvarsson 2003; Teixeira and Bernasconi 2008). In *R. nivalis*, X-ratio bias was observed in pollen grains before pollination using flow cytometry, although this was less strong than the sex ratio bias observed after pollination (Stehlik et al. 2007). This suggests that pollen drive may contribute at least in part to the observed sex ratio bias in this species, but the generality of this result remains uncertain, given the limited applicability of the flow-cytometry approach. Pollen drive contributes to pollen abortion in maize-teosinte hybrids through gametophytic hybrid incompatibility and affects maize domestication (Berube et al. 2024). Though meiotic drive can be a powerful evolutionary force, we have limited direct evidence on the mechanism, particularly in plants. Investigating meiotic drivers in pollen (hereafter 'pollen drive') can provide more insights on their genetics and evolutionary consequences.

The loci subject to pollen competition or pollen drive can be identified through transmission ratio distortion (TRD) in seeds and pollen grains. Pooled sequencing provides a cost-effective way to estimate genome-wide allele frequencies (Schlötterer et al. 2014); detecting TRD by sequencing large pools of gametes or offspring has been used in many plants and animals (reviewed in Fishman and McIntosh 2019). In *Arabidopsis*, bulk sequencing of pollen grains has identified segregation distortions related to reproductive isolation in intra- and (more commonly) inter-specific hybrids (Corbett-Detig et al. 2019; Condon et al. 2025). One of the distorting loci was

estimated to expand to an interval of 7.19 Mb in interspecific *Arabidopsis* hybrids (Corbett-Detig et al. 2019). In *Drosophila*, a centromere-linked meiotic driver was identified through pooled sequencing of embryos which showed decay in allele frequencies around the drive locus in a 50 Mb window due to recombination (Wei et al. 2017). Whole-genome pooled sequencing of pollen grains and developing seeds is a promising approach to scan for loci with transmission ratio bias following meiosis and pollen competition.

In this study, we used pooled sequences of mature pollen grains and developing seeds from crosses in a single large population to investigate evidence for variants subject to meiotic drive and gametophytic selection (pollen competition) in the dioecious annual *Rumex hastatulus* (Polygonaceae). Western populations of the obligately outcrossing and wind-pollinated *R. hastatulus* have an X/Y sex determination system, with homogametic (XX) females and heterogametic (XY) males (Smith 1963). Female-biased sex ratios were observed in natural populations (Conn and Blum 1981) due to pollen competition (Field et al. 2012), although the potential role for X-biased pollen drive has not yet been investigated in this species. Wind-pollinated plants like *R. hastatulus* have great opportunities for pollen competition due to their commonly uniovulate flowers and high pollen-to-ovule ratio (Friedman and Barrett 2009; Field et al. 2012), making *R. hastatulus* a promising system to test for segregation distortion during pollen competition. The abundant chromosomal rearrangements in the *R. hastatulus* genome (Sacchi et al. 2024) also offer candidates for testing TRD due to pollen drive. Additionally, in a genetic mapping population 48 sites were identified with transmission distortion from males in *R. hastatulus* (Rifkin et al. 2022), providing further support for standing variation for male fitness and potentially selection on male siring fitness detectable through TRD.

We performed independent crosses in *R. hastatulus* and sequenced leaf tissue, pools of mature pollen grains and developing seeds for each cross. Through a parentage analysis based on genetic relatedness among samples, we found that the genotypes of 'seeds' were largely the same as the maternal genotypes with no sign of the expected paternal genotype, likely reflecting the fact that the majority of 'seeds' collected were unfertilized ovules. Allele frequencies in pollen samples show heterogeneity among different male individuals and between chromosomes in some individuals. We tested for candidate loci showing deviation in allele frequencies in pollen compared to leaf tissues, but did not find strong signals of allele frequency distortion caused by meiotic drive. Haplotype divergence from the reference genome and repetitive regions in the

genome might affect the patterns of allele frequency heterogeneity more than pollen drive. Lastly, we found that sequence coverage on the Y-specific regions was significantly lower in pollen compared to male leaf tissues while comparing to the X-specific regions in most males or to autosomes in all males, which could suggest a signal of X ratio bias in pollen grains, where Y-bearing pollen has a disadvantage compared to X-bearing pollen.

## Results and Discussion

*Genetic relatedness among samples*

We performed 10 random crosses using male and female plants from different maternal families of *R. hastatulus* from a single large population and collected leaf, pollen and seed tissues for whole-genome sequencing (see Materials and Methods; Supplementary Table C1). We first tested the genetic relatedness of all DNA samples to confirm genetic correlations between seed samples and male and female leaves ('parents') from the same cross. The genetic relatedness matrix showed positive correlations among samples from the same maternal family and samples collected from the same individual, i.e., between seed and female leaf or pollen and male leaf. However, we found no genetic correlation between the seed and parental male leaf samples of the same cross (Supplementary Figure C2). Principal component analysis showed that samples from different maternal families were grouped into three clusters separated by PC1 and PC2 that explained 7.91% and 7.3% variance, respectively, suggesting some maternal families were more closely related than others (Supplementary Figure C3).

Our relatedness matrix suggested that our sequencing of pooled seed was in fact largely derived from maternal tissue rather than from the cross. To assess this further, we examined the genotypes of seeds where the parents were homozygous for different alleles, we expected the seed genotypes to be heterozygous. In most crosses, there were fewer than 0.3% of such sites in the seeds that were heterozygous, except for Cross 10 where there were 2% heterozygous sites in the seeds (Supplementary Table C2). In Cross 4, there were only 15 sites where the parents are homozygous for different alleles, due to a positive genetic correlation between the two parental leaf samples (55e male leaf × 31h female leaf), suggested by the genetic relatedness matrix (Supplementary Figure C2). One plausible explanation is that our pooled seed sequencing was heavily biased toward the maternal genotype due to a predominance of DNA from the maternal seed coat (entirely maternal) and the endosperm (two-thirds maternal), which could drive an

extreme bias in sequencing towards the maternal genotype. Alternatively, in an attempt to sequence early seeds that were less exposed to selection during seed development, we potentially collected largely unfertilized ovules. This would also explain the extreme signs of X sequencing bias in seeds from genomic coverage shown in Supplementary Figure C1c.

*Allele frequencies in pollen and leaf*

We investigated the signals of pollen drive on autosomes based on allele frequency distortions. Under Mendelian segregation, the allele frequency at heterozygous sites on autosomes is expected to be 50%. Any deviation of allele frequency between male leaf and pollen could be caused by meiotic drive in pollen grains. We tested for evidence of pollen drive by comparing the allelic depths in leaf and pollen from all males across the four autosomes of *R. hastatulus* (Figure 4.1, Supplementary Figures C4-C12). We filtered for heterozygous SNPs in males and removed sites with skewed allele frequencies in the leaf and used it for comparison (see Materials and Methods). Using one male as an example, we found heterogeneity in pollen allele frequencies across the genome (Figure 4.1a-c). On all autosomes, minor allele frequencies were consistently higher in leaf than pollen, driven by higher reference allele frequencies in pollen than leaf (Figure 4.1, Supplementary Figures C4-C12). We found particularly elevated reference allele frequencies in pollen relative to leaf on the right end of autosome A3 spanning a 100 Mb region (Figure 4.1b). Although this pattern could be reflective of a signal of pollen drive, it was consistently distorted across all crosses (Supplementary Figures C4-C12), making it more likely to be a technical artifact (see below).

Next, we performed Fisher's exact test at each heterozygous SNP to identify significant deviations in allelic depths between pollen and leaf, and calculated average *p*-values in sliding windows (see Materials and Methods). SNPs with lowest *p*-values will likely indicate allele frequency distortion, and we expected drive loci to show peaks in the Manhattan plot based on negative log values of *p*. There were varying patterns of allele frequency deviations across chromosomes and among different males (Figure 4.1, Supplementary Figures C4-C12). We found an increase in negative log values of *p* in the same region with elevated reference allele frequencies on A3 (Figure 4.1d), which could potentially indicate allele frequency distortions caused by pollen drive. However, why there were consistently higher reference allele frequencies in pollen than leaf remained unclear. Mapping bias toward the reference allele in short-read

alignment can lead to deviations from expected allele frequencies. Additionally, technical factors during DNA sample preparation might affect allele frequency; due to the low DNA concentration in leaf samples, PCR amplification was used during library preparation of leaf samples rather than pollen samples, which may bias allele frequency estimates. Lastly, different distributions of repetitive content and/or haplotype divergence from the reference genome likely contributed to the heterogeneity in allele frequency bias across the genome and among different individuals.
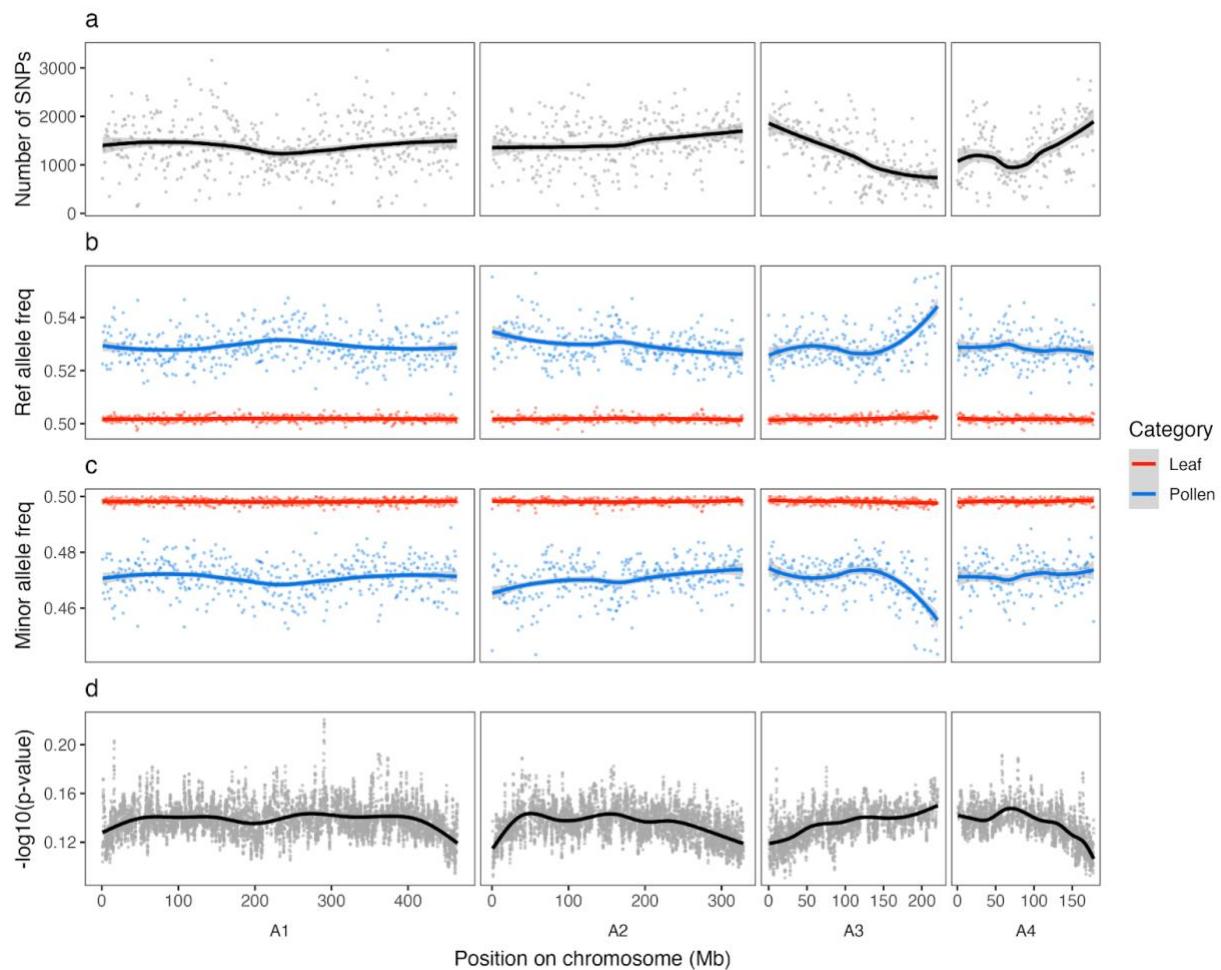


Figure 4.1. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 8. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.

## Genomic coverage on sex chromosomes

We examined signals of sex ratio bias due to sex chromosome drive using genomic coverage. If X-bearing pollen shows distortion over Y-bearing pollen, we would expect an X-biased ratio in pollen grains indicated by higher coverage on the X and lower coverage on the Y compared to leaves, using autosomal coverage as a reference point. We compared the number of mapped reads on X- or Y-specific regions to each autosome and all autosomes combined in pollen and male leaf using contingency tests, an odds ratio less than 1 indicates lower X or Y than autosomal coverage (Supplementary Table C3-C4). We found 8 of 10 males showed significantly lower coverage on Y than each autosome individually or all autosomes combined in pollen compared with leaves ($p$-value = 0, odds ratio between 0.85 – 0.98, Chi-squared test), potentially suggesting a preference for the X over the Y chromosome. However, the same males did not consistently show higher X coverage than autosomes, e.g., all males showed significantly lower X coverage than the second autosome (A2) in pollen compared with leaves ($p$-value < $10^{-16}$, odds ratio between 0.88 – 1). Significantly higher X coverage was found in at most half of the males when compared to individual autosomes or all autosomes combined ($p$-value < 0.05, odds ratio between 1 – 1.02). When directly comparing the coverage between the X and Y chromosomes, 8 of 10 males showed a significantly higher X coverage relative to the Y ($p$-value < $10^{-16}$, odds ratio between 1 – 1.15), providing some evidence of X segregation distortion.

Overall, the higher X-to-Y coverage ratio in pollen from most males (8 out of 10), along with the consistently lower Y coverage compared to autosomes, suggested a possible bias toward the X chromosome in mature pollen grains. However, the lack of a corresponding increase in X coverage relative to autosomes combined with the signs of DNA preparation biases from the allele frequency patterns make it unclear if there was truly X-biased distortion. One possible explanation for the discrepancy between X and Y coverage signals compared with autosomes is that the relatively high polymorphism levels on the X make it more difficult to accurately estimate the abundance of the X in pooled pollen, whereas the low Y polymorphism (Hough et al. 2017) yields a more accurate and consistent measure of relative coverage.

# Conclusion and Future Directions

Despite the technical challenges, our results laid a foundation for the lab's future investigations on pollen competition and pollen drive in plant species. We were motivated by the importance of testing for the genome-wide prevalence of standing variation for pollen competition and the success in other studies that tested for allele frequency distortions using large pools of gametes or progenies (e.g., Corbett-Detig et al. 2015, 2019; Bélanger et al. 2016; Wei et al. 2017; Condon et al. 2025). Collecting seeds to assess allele frequency distortions in progeny has several challenges. First, the seed coat and endosperm introduce extra maternal tissues into sequencing, which inflates the baseline allele frequency in seeds from a 50-50 ratio between maternal and paternal genetic material in embryos. While it is ideal to sequence embryos for testing transmission ratio distortion, seeds of wind-pollinated plants like *R. hastaulus* are usually small (less than 2 mm long when mature), and collecting large pools of embryos remains laborious and technically challenging. Second, understanding seed developmental stages is essential. Mature seeds will have gone through developmental selection, seed abortion, which if unaccounted for in sequencing experiments, will lead to bias in allele frequencies in mature seeds. A balance between maximizing the number of seeds to increase the statistical power of detecting allele frequency differences and the amount of labor and effort to collect the optimal developmental stage of early seeds is still challenging, especially in a system where we lack detailed knowledge on the developmental biology of seeds. The intensity of pollen competition can be increased by the quantity of pollen deposited on the stigma; insect-pollinated plant systems can potentially be good candidate systems to study the genetic architecture of pollen competition. Even with the biases introduced by non-embryo tissues in the seeds, by combining patterns of differential gene expression between pollen and other sporophytic tissues and gene functions, we can still further confirm the genomic regions subject to pollen competition. With the availability of parental genotypes, a likelihood ratio test that incorporates information on recombination rates will be a more powerful test on signals of distortion (Corbett-Detig et al. 2019; Condon et al. 2025). Lastly, testing whether candidate pollen drive loci overlap among different male plants will be particularly interesting for future studies, as this would be indicative of common alleles segregating within populations.

Other than standing variation in pollen competitive traits, pollen competition outcomes are also affected by the maternal control of the interaction between pistil that includes stigma, style and

ovule and pollen (Williams and Mazer 2016). Despite more technical difficulty in isolating the involvement of ovules during fertilization compared to pollen (Beaudry et al. 2020), both pre- and post-zygotic reproductive success controlled by the maternal genotype is worth further investigation. For example, longer styles create more room for pollen tubes to compete thus creates stronger pollen competition (Mazer et al. 2010; 2016). Pollen tube growth is also guided by molecules produced by the style before and after the pollen tube reaches the ovule (Higashiyama and Takeuchi 2015). The timing of stigma receptivity after pollen germination can also affect seed sets (Lankinen et al. 2016).

With limitations of our dataset, we did not find clear signals of pollen drive. This highlights the multiple biological and technical factors in shaping allele frequency patterns across the genome. For example, divergence from the reference genome and copy number variation (CNV) across genomic regions or individuals can cause reads to map incorrectly, introducing bias in allele frequency estimation (Schlötterer et al. 2014). Additionally, standardized methods during DNA sample preparation and sequencing will ensure an unbiased estimation of allele frequencies. Overall however, the lack of compelling evidence for pollen drive within a single population is in line with other studies showing more extensive signs of allelic distortion in between-population and between-species crosses in *Arabidopsis,* which was caused by genetic incompatibilities rather than meiotic drive (Corbett-Detig et al. 2019; Condon et al. 2025).

## Materials and Methods

### Plant materials

We collected open-pollinated seeds from a population of *R. hastatulus* in Rosebud, Texas, US (Pickup and Barrett 2013) and generated $F_1$ families by performing random crosses among families, as described in Chapters 2 and 3. We used seeds from 14 previous random crosses to generate $F_1$ plants: seeds were germinated on moist petri dishes at 4 °C and sprouts were transferred to soil after day 10 and repotted to 6-inch pots after day 28. We randomly paired one male and one female flowering individual from different families in a mini chamber for crossing, using a modified setup based on McGoey et al. (2017). We conducted 10 independent crosses in the glasshouse of the Earth Sciences Center, University of Toronto. Procedures to avoid pollen contamination among crosses included removing female inflorescences before they were placed in the mini chambers, thoroughly spraying both male and female plants with water mist to

remove residual pollen grains before crossing and placing a single female plant in a mini chamber as a control for pollen cross contamination. The pollen load on female flowers was maximized as most male plants were in peak flowering during crossing.

We collected leaf tissues and mature pollen grains from male and female plants before crossing. Pollen isolation methods were described in Chapter 2. We collected pools of early-staged seeds including aborted seeds from each cross 2 weeks after crossing, to prevent from collecting seeds that might have been fertilized by residual pollen grains introduced before crossing. Seeds were collected on multiple days until a minimum number of seeds were collected for each cross, multiple collections of seeds were merged during DNA isolation; the rough number of seeds collected per cross range from 100 to 230. All plant tissues were flash frozen in liquid nitrogen and stored in -80 degrees freezer before DNA isolation. We extracted DNA for all samples (male leaf, female leaf, pollen and seed) of each cross using Qiagen Plant Mini Kit following the standard protocol. Illumina libraries were prepared and sequenced at the Centre for Applied Genomics (Hospital for Sick Children), Toronto, Canada. Samples were sequenced on one lane of NovaSeq6000 S4 flowcell (PE150). All leaf DNA samples were sequenced with PCR during library preparation due to low DNA concentration of some samples. The coverage for leaf DNA samples were 10-15X, the coverage for seed and pollen DNA samples were at least 50X. There were a total of 40 DNA samples.

## Variant calling and filtering

We performed quality control of raw sequences using FastQC (Andrews 2010). We mapped all sequences to a phased genome assembly containing 4 autosomes, X and Y sex chromosomes (Sacchi et al., in prep) using bwa-mem2 (Vasimuddin et al. 2019), and sorted the BAM, added read groups, and removed duplicates using PicardTools (http://broadinstitute.github.io/picard/). We assessed the coverage of each sample across the genome using Qualimap BAM QC (Okonechnikov et al. 2016) and tinycov (https://github.com/cmdoret/tinycov). We counted the number of mapped reads for each chromosome using SAMTools (Danecek et al. 2021).

We then called variants jointly on all DNA samples with a ploidy of 2 using BCFTools mpileup (Danecek et al. 2021). We filtered for SNPs using the VCF with all samples to assess the genetic relatedness among samples: we filtered for bi-allelic SNPs on autosomes with a minimal mean read depth per sample between 10 and 200, a minimal phred-scaled quality score (QUAL) of 30,

and zero missing data for all samples using VCFTools (Danecek et al. 2011). We then removed SNPs with a minor allele frequency lower than 0.05 or with strong Hardy–Weinberg deviations ($p < 10_{-6}$) and performed linkage pruning using Plink (Chang et al. 2015). We performed principal component analysis (PCA) and calculated genetic relatedness matrix from linkage pruned sites using Plink (Chang et al. 2015).

To test the genotypes of seeds, we performed SNP filtering for each cross separately. For leaf samples, we filtered for bi-allelic SNPs with high genotype quality (GQ > 50) and a read depth smaller than twice the mean coverage on all autosomes of that sample. For seed samples, we filtered for sites with a read depth between 50 and twice the mean coverage of that sample. We filtered for SNPs where male and female leaves of the same cross were homozygous for different alleles, and kept the sites with zero missing data for male leaf, female leaf, and seed. Similarly, we performed SNP filtering for each pair of leaf and pollen samples from the same male individual to compare their allele frequencies. We filtered for sites with a read depth between 50 and twice the mean coverage of the sample for the pollen sample, and used the same filtering criteria for the leaf sample as mentioned above. We removed multi-allelic sites in pollen where the third and fourth allele have a combined allele frequency larger than 5%, based on allelic depths in the pileup file. We only kept heterozygous sites in both male leaf and pollen and removed sites where the reference allele frequency in the leaf is > 0.55 or < 0.45.

## *Allele frequency analysis*

To compare the allele frequencies on autosomes in male leaf and pollen, we calculated reference and minor allele frequency in sliding windows along each chromosome with a window size and step size of 1Mb. We only kept windows with a minimum of 50 SNPs. We performed Fisher's exact test on allelic depths of the reference and alternative allele at each site to test for allele frequency distortion. We then calculated the average of $p$-values in sliding windows with a window size of 1000 SNPs and a step size of 100 SNPs. To plot the average $p$-values along each chromosome, we used the mid position between the first and last SNP of each window as the position for each window. We added smooth lines of average $p$-values generated by the default smoothing function in R.

## Data availability

Raw sequencing reads were deposited to NCBI Genbank (BioProject PRJNA983258). The scripts used in this study are available at https://github.com/imengyuan/pollen_competition.

## References

Austen EJ, Weis AE. 2016a. The causes of selection on flowering time through male fitness in a hermaphroditic annual plant: flowering time and male fitness. *Evolution* 70:111–125.

Austen EJ, Weis AE. 2016b. Estimating selection through male fitness: three complementary methods illuminate the nature and causes of selection on flowering time. *Proc. R. Soc. B Biol. Sci.* 283:20152635.

Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI. 2020. Evolutionary genomics of plant gametophytic selection. *Plant Commun.* 1:100115.

Bélanger S, Clermont I, Esteves P, Belzile F. 2016. Extent and overlap of segregation distortion regions in 12 barley crosses determined via a Pool-GBS approach. *Theor. Appl. Genet.* 129:1393–1404.

Bernasconi G. 2003. Seed paternity in flowering plants: an evolutionary perspective. *Perspect. Plant Ecol. Evol. Syst.* 6:149–158.

Berube B, Ernst E, Cahn J, Roche B, de Santis Alves C, Lynn J, Scheben A, Grimanelli D, Siepel A, Ross-Ibarra J. 2024. Teosinte Pollen Drive guides maize diversification and domestication by RNAi. *Nature* 633:380–388.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:s13742-015-0047–0048.

Chibalina MV, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* 21:1475–1479.

Condon C, Carpentier F, Tabourin M, Wozniak N, Takou M, Blassiau C, Kumar V, Pietzenuk B, Habert R, De Meaux J, et al. 2025. Diverging *Arabidopsis* populations quickly accumulate pollen-acting genetic incompatibilities. *Evol. Lett.*:qraf013.

Conn JS, Blum U. 1981. Sex ratio of *Rumex hastatulus*: the effect of environmental factors and certation. *Evolution* 35:1108–1116.

Corbett-Detig R, Jacobs-Palmer E, Hartl D, Hoekstra H. 2015. Direct gamete sequencing reveals no evidence for segregation distortion in house mouse hybrids. *PLOS ONE* 10:e0131933.

Corbett-Detig R, Medina P, Frérot H, Blassiau C, Castric V. 2019. Bulk pollen sequencing reveals rapid evolution of segregation distortion in the male germline of *Arabidopsis hybrids*. *Evol. Lett.* 3:93–103.

Correns C. 1928. Bestimmung, vererbung and verteilung des geschlechtes bei den hoheren pflanzen. *Handbuch Vererbungsw* 2: 1–138.

Crow JF, Kimura M. 1965. Evolution in sexual and asexual populations. *Am. Nat.* 99:439–450.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.

Delph LF. 2019. Pollen competition is the mechanism underlying a variety of evolutionary phenomena in dioecious plants. *New Phytol.*:nph.15868.

Field DL, Pickup M, Barrett SCH. 2012. The influence of pollination intensity on fertilization success, progeny sex ratio, and fitness in a wind-pollinated, dioecious plant. *Int. J. Plant Sci.* 173:184–191.

Fishman L, McIntosh M. 2019. Standard deviations: the biological bases of transmission ratio distortion. *Annu. Rev. Genet.* 53:annurev-genet-112618-043905.

Friedman J, Barrett SC. 2009. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Ann. Bot.* 103:1515–1527.

Higashiyama T, Takeuchi H. 2015. The mechanism and key molecules involved in pollen tube guidance. *Annu. Rev. Plant Biol.* 66:393–413.

Hormaza JI, Herrero M. 1992. Pollen selection. *Theor. Appl. Genet.* 83:663–672.

Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. 2014. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proc. Natl. Acad. Sci.* 111:7713–7718.

Hough J, Wang W, Barrett SCH, Wright SI. 2017. Hill-Robertson interference reduces genetic diversity on a young plant Y-chromosome. *Genetics* 207:685–695.

Immler S. 2019. Haploid selection in "diploid" organisms. *Annu. Rev. Ecol. Evol. Syst.* 50:219–236.

Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? *Nature* 351:314–315.

Lankinen Å, Smith HG, Andersson S, Madjidian JA. 2016. Selection on pollen and pistil traits during pollen competition is affected by both sexual conflict and mixed mating in a self-compatible herb. *Am. J. Bot.* 103:541–552.

Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, Johannesson H, Knief U, Kokko H, Larracuente AM, et al. 2016. The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* 31:315–326.

Lloyd DG. 1974. Female-predominant sex ratios in angiosperms. *Heredity* 32:35–44.

Mazer SJ, Dudley LS, Hove AA, Emms SK, Verhoeven AS. 2010. Physiological performance in Clarkia sister taxa with contrasting mating systems: do early-flowering autogamous taxa avoid water stress relative to their pollinator-dependent counterparts? *Int. J. Plant Sci.* 171:1029–1047.

Mazer SJ, Moghaddasi A, Bello AK, Hove AA. 2016. Winning in style: Longer styles receive more pollen, but style length does not affect pollen attrition in wild Clarkia populations. *Am. J. Bot.* 103:408–422.

McGoey BV, Janik R, Stinchcombe JR. 2017. Individual chambers for controlling crosses in wind-pollinated plants. *Methods Ecol. Evol.* 8:887–891.

Monnahan PJ, Colicchio J, Fishman L, Macdonald SJ, Kelly JK. 2021. Predicting evolutionary change at the DNA level in a natural Mimulus population. *PLOS Genet.* 17:e1008945.

Mulcahy DL, Mulcahy GB. 1987. The effects of pollen competition. *Am. Sci.* 75:44–50.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294.

Pickup M, Barrett SCH. 2013. The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. *Ecol. Evol.* 3:629–639.

Presgraves D. 2009. Drive and sperm: the evolution and genetics of male meiotic drive. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm Biology. London: Academic Press. p. 471–506.

Rifkin JL, Hnatovska S, Yuan M, Sacchi BM, Choudhury BI, Gong Y, Rastas P, Barrett SCH, Wright SI. 2022. Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 377:20210226.

Rychlewski J, Zarzycki K. 1975. Sex ratio in seeds of *Rumex acetosa* L. as a result of sparse or abundant pollination. *Genet. Pol.* 17.

Rychlewski J, Zarzycki K. 1986. Genetical and ecological mechanisms regulating the sex ratio in populations of *Rumex thyrsiflorus* Fingerh.(Polygonaceae). *VEROEFF GEOBOT INST ETH STIFT RUEBEL* 87:132–140.

Sacchi B, Humphries Z, Kružlicová J, Bodláková M, Pyne C, Choudhury BI, Gong Y, Bačovský V, Hobza R, Barrett SCH, et al. 2024. Phased assembly of neo-sex chromosomes reveals extensive Y degeneration and rapid genome evolution in *Rumex hastatulus*. *Mol. Biol. Evol.* 41:msae074.

Sandler G, Beaudry FEG, Barrett SCH, Wright SI. 2018. The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants. *Evol. Lett.* 2:368–377.

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15:749–763.

Scott MF, Otto SP. 2017. Haploid selection favours suppressed recombination between sex chromosomes despite causing biased sex ratios. *Genetics*:genetics.300062.2017.

Smith BW. 1963. The mechanism of sex determination in *Rumex hastatulus*. *Genetics* 48:1265–1288.

Stehlik I, Barrett SC. 2006. Pollination intensity influences sex ratios in dioecious *Rumex nivalis*, a wind-pollinated plant. *Evolution* 60:1207–1214.

Stehlik I, Kron P, Barrett SCH, Husband BC. 2007. Sexing pollen reveals female bias in a dioecious plant. *New Phytol.* 175:185–194.

Taylor DR. 1996. Parental expenditure and offspring sex ratios in the dioecious plant *Silene alba* (= *Silene latifolia*). *Am. Nat.* 147:870–879.

Taylor DR. 1999. Genetics of sex ratio variation among natural populations of a dioecious plant. *Evolution* 53:55–62.

Taylor DR, Ingvarsson PK. 2003. Common features of segregation distortion in plants and animals. *Genetica* 117:27–35.

Teixeira S, Bernasconi G. 2008. Effects of inbred/outbred crosses on progeny sex ratio in *Silene latifolia* (Caryophyllaceae). *New Phytol.* 178:448–456.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). p. 314–324.

Walsh NE, Charlesworth D. 1992. Evolutionary interpretations of differences in pollen tube growth rates. *Q. Rev. Biol.* 67:19–37.

Wei KH-C, Reddy HM, Rathnam C, Lee J, Lin D, Ji S, Mason JM, Clark AG, Barbash DA. 2017. A pooled sequencing approach identifies a candidate meiotic driver in *Drosophila*. *Genetics* 206:451–465.

Williams JH, Mazer SJ. 2016. Pollen—tiny and ephemeral but not forgotten: new ideas on their ecology and evolution. *Am. J. Bot.* 103:365–374.

# Chapter 5: Concluding remarks

In my thesis I explored several questions on the evolutionary conflict between plant life stages through population genomics and experimental crosses, motivated by the consequences of the alternation between diploid and haploid life stages. Below I briefly summarize the implications of my research to the broader field of evolutionary genetics and plant biology and discuss the limitations of current tests for selection and conflict, and directions for future research.

In Chapter 2, I tested the potential for ongoing conflict between life stages and sexes in the dioecious angiosperm *Rumex hastatulus* based on *cis*-regulatory variation. In Chapter 3, I followed up with population genomic tests of balancing selection due to the intralocus conflict between life stages in the angiosperm *R. hastatulus* and the moss *Ceratodon purpureus*. Based on the results of these two chapters, I found limited evidence for ongoing conflict between the diploid and haploid phase of the plant lifecycle, given the evidence for low genetic correlations in expression in Chapter 2, and no sign of genome-wide balancing selection in Chapter 3. My results highlight the challenges in detecting balancing selection generated by life-stage conflict, and that concordant selection might be widespread between life stages and is as important as antagonistic selection in shaping gene expression variation.

In Chapter 2, the positively correlated effects of *cis*-regulatory variation between sexes and life stages in the dioecious *Rumex hastatulus* suggested potential scope for both sexual and life-stage conflict. Plant systems offer many opportunities to test the potential and patterns of sexual conflict within diploid or haploid stages and conflict co-occurring between sexes and life-stages (e.g., between haploid male and diploid female or haploid female and diploid male). Testing selective forces on regulatory variation using association mapping helps answer fundamental questions such as the maintenance of genetic variation in quantitative traits (Josephs et al. 2017). While we did not find a pattern of purifying selection on cis-regulatory variation in *R. hastatulus*, follow-up analyses, such as those comparing the burden of rare alleles in leaf vs pollen may provide a more powerful approach to detect purifying selection (Uzunović et al. 2019).

As suggested in Chapter 2, the extent for potential conflict between life stages was smaller than sexual conflict in *R. hastatulus* based on our sampling of tissues. Despite a smaller potential, genes with overlapping expression between gametophytic and sporophytic tissues might still show ongoing conflict and be enriched for signals of balancing selection. In Chapter 3, I tested

for signals of balancing selection due to intralocus conflict between haploid and diploid life stages in *R. hastatulus* and *C. purpureus*. Current empirical approaches for studying balancing selection can be divided into two categories: the first one is knowing the biology of traits under balancing selection then test its genomic signal and genetic basis, the second one is identifying candidate loci under balancing selection using genomic signatures without prior knowledge of specific traits under selection then examine the functional annotation of candidate loci (Ruzicka et al. 2025). The first approach could be biased towards identifying a relatively small number of loci with large effects, while the second approach sometimes lacks power to detect true targets of balancing selection and to differentiate different mechanisms causing it. In Chapter 3, I used the second approach with both traditionally used diversity statistics and model-based tests in both *R hastatulus* and *C. purpureus*. Even though I did not find compelling signal of life-stage conflict and the genome-wide patterns are more consistent with synergistic pleiotropy in both species, balancing selection due to intralocus conflict between life stages or other processes is still quite possible on specific target loci and will require a deeper understanding of the biological or environmental factors driving different selective pressures and causing conflict, along with their genetic basis. Additionally, limited knowledge of gene annotation and gene functions in non-model systems makes it difficult to interpret functional enrichment results of candidate loci or loci linked to the true targets of selection. Understanding the prevalence and biological meaning of balancing selection will require continuing theoretical and empirical efforts. Lastly, there is a less clear prediction on the signal of interlocus conflict, e.g., due to co-evolution of different pollen or sperm competitive alleles (Clark 2002), but its possibility should be further explored in plant systems.

With the notable difficulty of measuring fitness in natural populations, alternative methods have been used in my thesis to study selection and to examine the potential scope for conflict. Diversity statistics (e.g., nucleotide diversity, Tajima's $D$) have been a useful tool in testing the signatures of selection, however, they suffer from biases introduced by demographic history. In study systems where knowledge of demographic history is unavailable, making inference on selection based on diversity statistics alone needs to be done with caution. High mutation rate elevates polymorphism levels, which means that model-based tests of selection that are robust to mutation rate variation will also be essential. In Chapters 2 and 3 I used gene expression as a focal trait to test the potential for antagonistic selection under the conflict between sexes or life

stages. Gene expression is an important molecular trait underlying phenotypic variation, it is accessible to study but lacks the direct link to fitness in non-model systems. With the lack of fitness data, only indirect tests of antagonistic selection can be performed, attempts to measure male and female fitness in dioecious plant systems will be important in future tests of selection between sexes or life stages (e.g., in Chapter 4, I attempted to identify loci affecting male siring success during pollen competition).

A more comprehensive tissue sampling in the angiosperm *R. hastatulus* will help provide a clearer interpretation of the findings. In Chapter 2, using pollen and leaf alone may not be representative of the two life stages, both tissues have very specialized functions and the ability to draw general conclusions on haploid and diploid life stages is limited; other gametophytic (e.g., ovule) and sporophytic tissues (e.g., roots, stem) need to be sampled. In Chapter 3, the limited tissue sampling in *R. hastatulus* made it difficult to control for all the factors affecting selection such as expression breath. In the moss *C. purpureus*, using the whole plant of each life stage offers a direct comparison that bypasses the need to sample different tissue types. However, it remains unclear why gametophytic-specific genes do not experience stronger purifying selection than sporophyte-specific genes in *C. purpureus*. Different cell types can have different expression patterns and experience different selective pressures within a single tissue; future studies will benefit from comparing cell types and using single cell transcriptomics (Price et al. 2022; Darolti and Mank 2023), e.g. in pollen and ovule. Other model systems with population genomic data and comprehensive transcriptomic data across tissues and life stages will be suitable to test the signal of intralocus conflict between life stages or sexes (e.g. in the monoecious maize).

In Chapter 4, characterizing standing variation for pollen meiotic drive and pollen competition is vital to understand the processes affecting pre-zygotic fertilization success. Despite many technical challenges, identifying genetic variation for pollen competition, and testing its signatures of selection and potential roles in the interactions between pollen and ovule or the sporophytic stage is valuable for future research. It would also be interesting to test selection on pollen competitive ability in comparisons of mating systems. Self-fertilization reduces the potential of pollen competition, due to the increase in genetic similarity among pollen grains, and selection for pollen competitive ability is relaxed (Mazer et al. 2010; Peters and Weis 2019). Wind-pollication occurs in both angiosperm and gymnosperm species, the potential for pollen

competition is higher in angiosperms than gymnosperms in terms of length of travel for pollen tubes, pollen-ovule ratio, and timing of pollen deposition (Mulcahy and Mulcahy 1987). Whether the lower pollen competition intensity in gymnosperms suggest lower or higher potential for conflict between gametophytes and sporophytes is of interest for future work. Even though my results were inconclusive, testing sex ratio bias in gametophytes/gametes using genomic coverage data is useful for future studies on different mechanisms of sex ratio bias in plants and animals (pollen/sperm competition, male and female meiotic drive).

Overall, my research highlights the opportunities of using the diverse life cycles, life histories and mating systems of plants to study fundamental questions in evolutionary genetics and plant biology. In Chapter 3, my results suggested different extent of gametophytic expression between the angiosperm and the moss. Future studies should test how the varying complexity and length of gametophytic and sporophytic stages affect the expression patterns between life stages (and potential for conflict) in a more comparative context and include other major clades of land plants in addition to angiosperms and bryophytes (lycophytes, ferns, gymnosperms) (e.g., Sorojsrisom 2025). Comparing expression and gene regulatory networks between life stages will shed light on the adaptation of gametophytes and sporophytes and land plant evolution (Nishiyama et al. 2003; Szövényi et al. 2011). In bryophytes like *C. purpureus*, gametophytic selection is key to the evolution of haploid sex chromosomes (Charlesworth 2025). Sex-linked genes with sporophyte-specific expression would be important in testing the parent-offspring conflict between sporophytes and female gametophytes. With the increasing availability of genomic data, testing the pleiotropy between plant life stages and sexes in different plant lineages and how it affects the maintenance of genetic variation should be a main focus in future studies.

# Reference

Charlesworth D. 2025. Sex chromosome evolution in haploid plants: microchromosomes, disappearing chromosomes, and giant chromosomes. Proc. Natl. Acad. Sci. U.S.A. 122:e2425050122.

Clark AG. 2002. Sperm competition and the maintenance of polymorphism. Heredity 88:148–153.

Darolti I, Mank JE. 2023. Sex-biased gene expression at single-cell resolution: cause and consequence of sexual dimorphism. Evol. Lett. 7:148–156.

Josephs EB, Stinchcombe JR, Wright SI. 2017. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? New Phytol. 214:21–33.

Mazer SJ, Hove AA, Miller BS, Barbet-Massin M. 2010. The joint evolution of mating system and pollen performance: Predictions regarding male gametophytic evolution in selfers vs. outcrossers. Perspect. Plant Ecol. Evol. Syst. 12:31–41.

Mulcahy DL, Mulcahy GB. 1987. The effects of pollen competition. Am. Sci. 75:44–50.

Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, et al. 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. Proc. Natl. Acad. Sci. 100:8007–8012.

Peters MAE, Weis AE. 2019. Isolation by phenology synergizes isolation by distance across a continuous landscape. New Phytol. 224:1215–1228.

Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022. Detecting signatures of selection on gene expression. Nat. Ecol. Evol. 6:1035–1045.

Ruzicka F, Zwoinska MK, Goedert D, Kokko H, Richter X-YL, Moodie IR, Nilén S, Olito C, Svensson EI, Czuppon P, et al. 2025. A century of theories of balancing selection. Biorxiv:2025.02.12.637871.

Sorojsrisom ES. 2025. It's not just a phase: development, evolutionary history, and consequences of the free-living gametophyte phase in the life cycles of ferns [dissertation]. [New York (NY)]: Columbia University.

Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ. 2011. Generation-biased gene expression in a bryophyte model system. Mol. Biol. Evol. 28:803–812.

Uzunović J, Josephs EB, Stinchcombe JR, Wright SI. 2019. Transposable elements are important contributors to standing variation in gene expression in *Capsella grandiflora*. Mol. Biol. Evol. 36:1734–1745.

# Appendix A: Chapter 2 Supplementary Materials

## Chapter 2 Supplementary Tables

Table A1. DNA and female leaf RNA samples of *Rumex hastatulus* used in this study.

| Sample | Number of Reads | Tissue | Sex | Description | Accession | SRA |
|---|---|---|---|---|---|---|
| 10aMLD | 70,401,053 | leaf | male | DNA | SAMN20804082 | SRR15988589 |
| 10bFLD | 73,982,852 | leaf | female | DNA | SAMN20803999 | SRR15988723 |
| 11aMLD | 65,271,710 | leaf | male | DNA | SAMN20804083 | SRR15988588 |
| 11bFLD | 71,665,827 | leaf | female | DNA | SAMN20804000 | SRR15988722 |
| 12cFLD | 81,704,991 | leaf | female | DNA | SAMN20804001 | SRR15988659 |
| 12fMLD | 71,585,739 | leaf | male | DNA | SAMN20804084 | SRR15988587 |
| 13aMLD | 75,833,953 | leaf | male | DNA | SAMN20804085 | SRR15988586 |
| 13bFLD | 66,739,339 | leaf | female | DNA | SAMN20804002 | SRR15988648 |
| 14eMLD | 78,433,350 | leaf | male | DNA | SAMN20804086 | SRR15988585 |
| 15aFLD | 88,888,320 | leaf | female | DNA | SAMN20804003 | SRR15988637 |
| 15dMLD | 79,907,220 | leaf | male | DNA | SAMN20804087 | SRR15988584 |
| 16aFLD | 70,985,471 | leaf | female | DNA | SAMN20804004 | SRR15988626 |
| 16bMLD | 80,082,079 | leaf | male | DNA | SAMN20804088 | SRR15988583 |
| 17aFLD | 73,331,351 | leaf | female | DNA | SAMN20804005 | SRR15988615 |
| 17bMLD | 74,477,143 | leaf | male | DNA | SAMN20804089 | SRR15988581 |
| 18aFLD | 67,094,643 | leaf | female | DNA | SAMN20804006 | SRR15988604 |
| 18cMLD | 69,744,303 | leaf | male | DNA | SAMN20804090 | SRR15988580 |
| 19aMLD | 75,266,775 | leaf | male | DNA | SAMN20804091 | SRR15988579 |
| 19eFLD | 68,872,081 | leaf | female | DNA | SAMN20804007 | SRR15988593 |
| 1dFLD | 69,906,108 | leaf | female | DNA | SAMN20804065 | SRR15988608 |
| 1fMLD | 73,525,903 | leaf | male | DNA | SAMN20804066 | SRR15988607 |
| 20aMLD | 73,636,846 | leaf | male | DNA | SAMN20804092 | SRR15988578 |
| 20bFLD | 80,301,917 | leaf | female | DNA | SAMN20804008 | SRR15988582 |
| 22dMLD | 88,166,656 | leaf | male | DNA | SAMN20804093 | SRR15988577 |
| 22eFLD | 74,303,943 | leaf | female | DNA | SAMN20804009 | SRR15988721 |
| 23aMLD | 72,946,087 | leaf | male | DNA | SAMN20804094 | SRR15988576 |
| 23fFLD | 72,494,184 | leaf | female | DNA | SAMN20804010 | SRR15988710 |
| 24aFLD | 71,353,332 | leaf | female | DNA | SAMN20804011 | SRR15988699 |
| 24fMLD | 166,032 | leaf | male | DNA | SAMN20804095 | SRR15988575 |
| 25aFLD | 86,107,323 | leaf | female | DNA | SAMN20804012 | SRR15988688 |
| 25dMLD | 76,424,225 | leaf | | DNA | SAMN20804096 | SRR15988574 |
| 26fMLD | 68,644,837 | leaf | male | DNA | SAMN20804097 | SRR15988573 |
| 27cFLD | 67,319,284 | leaf | female | DNA | SAMN20804013 | SRR15988677 |
| 27eMLD | 78,769,997 | leaf | male | DNA | SAMN20804098 | SRR15988572 |
| 28cFLD | 64,585,743 | leaf | female | DNA | SAMN20804014 | SRR15988666 |
| 28dMLD | 83,264,313 | leaf | male | DNA | SAMN20804099 | SRR15988720 |
| 29cFLD | 73,026,787 | leaf | female | DNA | SAMN20804015 | SRR15988663 |
| 29eMLD | 84,451,770 | leaf | male | DNA | SAMN20804100 | SRR15988719 |

| 2bMLD | 65,064,544 | leaf | male | DNA | SAMN20804067 | SRR15988606 |
|---|---|---|---|---|---|---|
| 2fFLD | 85,713,751 | leaf | female | DNA | SAMN20804068 | SRR15988605 |
| 30bFLD | 78,606,582 | leaf | female | DNA | SAMN20804016 | SRR15988662 |
| 30fMLD | 76,760,991 | leaf | male | DNA | SAMN20804101 | SRR15988718 |
| 31dFLD | 85,527,734 | leaf | female | DNA | SAMN20804017 | SRR15988661 |
| 31eMLD | 71,748,177 | leaf | male | DNA | SAMN20804102 | SRR15988717 |
| 32aMLD | 68,254,501 | leaf | male | DNA | SAMN20804103 | SRR15988716 |
| 32fFLD | 70,770,913 | leaf | female | DNA | SAMN20804018 | SRR15988660 |
| 33aMLD | 80,970,784 | leaf | male | DNA | SAMN20804104 | SRR15988715 |
| 33fFLD | 75,863,074 | leaf | female | DNA | SAMN20804019 | SRR15988658 |
| 34cMLD | 83,119,055 | leaf | male | DNA | SAMN20804105 | SRR15988714 |
| 34fFLD | 75,251,572 | leaf | female | DNA | SAMN20804020 | SRR15988657 |
| 35aMLD | 85,800,800 | leaf | male | DNA | SAMN20804106 | SRR15988713 |
| 35bFLD | 70,994,851 | leaf | female | DNA | SAMN20804021 | SRR15988656 |
| 36cMLD | 66,666,365 | leaf | male | DNA | SAMN20804107 | SRR15988712 |
| 36dFLD | 64,028,115 | leaf | female | DNA | SAMN20804022 | SRR15988655 |
| 37fMLD | 78,212,236 | leaf | male | DNA | SAMN20804108 | SRR15988711 |
| 38bMLD | 79,018,641 | leaf | male | DNA | SAMN20804109 | SRR15988709 |
| 38dFLD | 80,471,885 | leaf | female | DNA | SAMN20804023 | SRR15988654 |
| 39dFLD | 83,697,285 | leaf | female | DNA | SAMN20804024 | SRR15988653 |
| 39fMLD | 79,253,753 | leaf | male | DNA | SAMN20804110 | SRR15988708 |
| 3bFLD | 69,826,322 | leaf | female | DNA | SAMN20804069 | SRR15988603 |
| 3dMLD | 73,239,617 | leaf | male | DNA | SAMN20804070 | SRR15988602 |
| 40aFLD | 71,055,865 | leaf | female | DNA | SAMN20804025 | SRR15988652 |
| 40bMLD | 80,806,772 | leaf | male | DNA | SAMN20804111 | SRR15988707 |
| 41cFLD | 67,055,870 | leaf | female | DNA | SAMN20804026 | SRR15988651 |
| 41eMLD | 81,412,161 | leaf | male | DNA | SAMN20804112 | SRR15988706 |
| 42cMLD | 71,063,053 | leaf | male | DNA | SAMN20804113 | SRR15988705 |
| 42dFLD | 65,953,316 | leaf | female | DNA | SAMN20804027 | SRR15988650 |
| 43aFLD | 71,889,822 | leaf | female | DNA | SAMN20804028 | SRR15988649 |
| 43bMLD | 64,512,861 | leaf | male | DNA | SAMN20804114 | SRR15988704 |
| 44bMLD | 67,031,508 | leaf | male | DNA | SAMN20804115 | SRR15988703 |
| 44fFLD | 84,828,356 | leaf | female | DNA | SAMN20804029 | SRR15988647 |
| 45dMLD | 78,384,599 | leaf | male | DNA | SAMN20804116 | SRR15988702 |
| 46aFLD | 77,530,318 | leaf | female | DNA | SAMN20804030 | SRR15988646 |
| 46eMLD | 73,762,147 | leaf | male | DNA | SAMN20804117 | SRR15988701 |
| 47dMLD | 69,636,851 | leaf | male | DNA | SAMN20804118 | SRR15988700 |
| 47fFLD | 80,761,342 | leaf | female | DNA | SAMN20804031 | SRR15988645 |
| 48aMLD | 65,604,816 | leaf | male | DNA | SAMN20804119 | SRR15988698 |
| 48cFLD | 79,641,895 | leaf | female | DNA | SAMN20804032 | SRR15988644 |
| 49bFLD | 75,604,714 | leaf | female | DNA | SAMN20804033 | SRR15988643 |
| 49fMLD | 81,648,661 | leaf | male | DNA | SAMN20804120 | SRR15988697 |
| 4aMLD | 64,169,782 | leaf | male | DNA | SAMN20804071 | SRR15988601 |
| 4cFLD | 79,098,390 | leaf | female | DNA | SAMN20804072 | SRR15988600 |
| 50aFLD | 72,316,775 | leaf | female | DNA | SAMN20804034 | SRR15988642 |
| 50bMLD | 90,477,716 | leaf | male | DNA | SAMN20804121 | SRR15988696 |

| 51dMLD | 68,237,786 | leaf | male | DNA | SAMN20804122 | SRR15988695 |
|--------|-----------|------|------|-----|--------------|-------------|
| 51fFLD | 66,627,361 | leaf | female | DNA | SAMN20804035 | SRR15988641 |
| 52cFLD | 81,400,871 | leaf | female | DNA | SAMN20804036 | SRR15988640 |
| 52eMLD | 73,962,470 | leaf | male | DNA | SAMN20804123 | SRR15988694 |
| 53bMLD | 73,753,527 | leaf | male | DNA | SAMN20804124 | SRR15988693 |
| 53cFLD | 64,980,628 | leaf | female | DNA | SAMN20804037 | SRR15988639 |
| 54aFLD | 75,676,303 | leaf | female | DNA | SAMN20804038 | SRR15988638 |
| 54bMLD | 80,597,105 | leaf | male | DNA | SAMN20804125 | SRR15988692 |
| 55cFLD | 72,787,257 | leaf | female | DNA | SAMN20804039 | SRR15988636 |
| 55dMLD | 69,886,810 | leaf | male | DNA | SAMN20804126 | SRR15988691 |
| 56aFLD | 72,156,350 | leaf | female | DNA | SAMN20804040 | SRR15988635 |
| 56eMLD | 68,906,425 | leaf | male | DNA | SAMN20804127 | SRR15988690 |
| 57aMLD | 79,397,151 | leaf | male | DNA | SAMN20804128 | SRR15988689 |
| 57cFLD | 68,494,348 | leaf | female | DNA | SAMN20804041 | SRR15988634 |
| 58bFLD | 66,903,088 | leaf | female | DNA | SAMN20804042 | SRR15988633 |
| 58fMLD | 68,359,360 | leaf | male | DNA | SAMN20804129 | SRR15988687 |
| 59aFLD | 80,184,212 | leaf | female | DNA | SAMN20804043 | SRR15988632 |
| 59cMLD | 76,941,321 | leaf | male | DNA | SAMN20804130 | SRR15988686 |
| 5aMLD | 77,337,033 | leaf | male | DNA | SAMN20804073 | SRR15988599 |
| 5dFLD | 76,449,089 | leaf | female | DNA | SAMN20804074 | SRR15988598 |
| 60cFLD | 76,452,703 | leaf | female | DNA | SAMN20804044 | SRR15988631 |
| 60fMLD | 78,040,195 | leaf | male | DNA | SAMN20804131 | SRR15988685 |
| 61cMLD | 66,083,000 | leaf | male | DNA | SAMN20804132 | SRR15988684 |
| 61fFLD | 70,809,076 | leaf | female | DNA | SAMN20804045 | SRR15988630 |
| 62bMLD | 76,186,458 | leaf | male | DNA | SAMN20804133 | SRR15988683 |
| 62dFLD | 78,244,622 | leaf | female | DNA | SAMN20804046 | SRR15988629 |
| 63aMLD | 68,933,353 | leaf | male | DNA | SAMN20804134 | SRR15988682 |
| 63fFLD | 64,280,172 | leaf | female | DNA | SAMN20804047 | SRR15988628 |
| 64aFLD | 65,552,643 | leaf | female | DNA | SAMN20804048 | SRR15988627 |
| 65cMLD | 75,476,485 | leaf | male | DNA | SAMN20804135 | SRR15988681 |
| 65dFLD | 66,799,342 | leaf | female | DNA | SAMN20804049 | SRR15988625 |
| 66aFLD | 74,662,391 | leaf | female | DNA | SAMN20804050 | SRR15988624 |
| 66dMLD | 71,625,543 | leaf | male | DNA | SAMN20804136 | SRR15988680 |
| 67fMLD | 82,513,664 | leaf | male | DNA | SAMN20804137 | SRR15988679 |
| 68cFLD | 74,183,480 | leaf | female | DNA | SAMN20804051 | SRR15988623 |
| 68eMLD | 73,487,331 | leaf | male | DNA | SAMN20804138 | SRR15988678 |
| 69aFLD | 74,988,869 | leaf | female | DNA | SAMN20804052 | SRR15988622 |
| 69dMLD | 72,408,399 | leaf | male | DNA | SAMN20804139 | SRR15988676 |
| 6aFLD | 73,336,537 | leaf | female | DNA | SAMN20804075 | SRR15988597 |
| 6dMLD | 87,373,454 | leaf | male | DNA | SAMN20804076 | SRR15988596 |
| 70bMLD | 76,568,929 | leaf | male | DNA | SAMN20804140 | SRR15988675 |
| 70fFLD | 70,692,499 | leaf | female | DNA | SAMN20804053 | SRR15988621 |
| 71aFLD | 78,880,098 | leaf | female | DNA | SAMN20804054 | SRR15988620 |
| 71eMLD | 72,065,237 | leaf | male | DNA | SAMN20804141 | SRR15988674 |
| 72dFLD | 72,557,216 | leaf | female | DNA | SAMN20804055 | SRR15988619 |
| 73aMLD | 74,543,572 | leaf | male | DNA | SAMN20804142 | SRR15988673 |

| 73dFLD | 79,203,830 | leaf | female | DNA | SAMN20804056 | SRR15988618 |
|--------|------------|------|--------|-----|--------------|-------------|
| 74bMLD | 86,718,202 | leaf | male | DNA | SAMN20804143 | SRR15988672 |
| 74eFLD | 75,119,886 | leaf | female | DNA | SAMN20804057 | SRR15988617 |
| 75dFLD | 86,526,559 | leaf | female | DNA | SAMN20804058 | SRR15988616 |
| 75fMLD | 78,594,151 | leaf | male | DNA | SAMN20804144 | SRR15988671 |
| 76bMLD | 69,610,876 | leaf | male | DNA | SAMN20804145 | SRR15988670 |
| 76dFLD | 83,522,622 | leaf | female | DNA | SAMN20804059 | SRR15988614 |
| 77bMLD | 70,470,899 | leaf | male | DNA | SAMN20804146 | SRR15988669 |
| 77fFLD | 70,831,968 | leaf | female | DNA | SAMN20804060 | SRR15988613 |
| 78aFLD | 67,026,638 | leaf | female | DNA | SAMN20804061 | SRR15988612 |
| 78eMLD | 76,837,646 | leaf | male | DNA | SAMN20804147 | SRR15988668 |
| 79aFLD | 80,566,146 | leaf | female | DNA | SAMN20804062 | SRR15988611 |
| 79bMLD | 68,616,128 | leaf | male | DNA | SAMN20804148 | SRR15988667 |
| 7bMLD | 77,227,398 | leaf | male | DNA | SAMN20804077 | SRR15988595 |
| 7cFLD | 73,958,129 | leaf | female | DNA | SAMN20804078 | SRR15988594 |
| 80aMLD | 87,038,758 | leaf | male | DNA | SAMN20804149 | SRR15988665 |
| 80bFLD | 86,778,222 | leaf | female | DNA | SAMN20804063 | SRR15988610 |
| 81cFLD | 75,402,838 | leaf | female | DNA | SAMN20804064 | SRR15988609 |
| 81dMLD | 75,939,039 | leaf | male | DNA | SAMN20804150 | SRR15988664 |
| 8aFLD | 80,953,696 | leaf | female | DNA | SAMN20804079 | SRR15988592 |
| 9bMLD | 82,607,803 | leaf | male | DNA | SAMN20804080 | SRR15988591 |
| 9cFLD | 94,662,110 | leaf | female | DNA | SAMN20804081 | SRR15988590 |
| 10bFLR | 23,457,246 | leaf | female | RNA | SAMN20803187 | SRR15881870 |
| 11bFLR | 21,849,240 | leaf | female | RNA | SAMN20803188 | SRR15881869 |
| 12cFLR | 23,553,440 | leaf | female | RNA | SAMN20803189 | SRR15881758 |
| 13bFLR | 20,903,566 | leaf | female | RNA | SAMN20803190 | SRR15881717 |
| 14bFLR | 32,889,641 | leaf | female | RNA | SAMN20803191 | SRR15881706 |
| 15aFLR | 24,918,133 | leaf | female | RNA | SAMN20803192 | SRR15881695 |
| 16aFLR | 25,875,155 | leaf | female | RNA | SAMN20803193 | SRR15881684 |
| 17aFLR | 21,289,491 | leaf | female | RNA | SAMN20803194 | SRR15881673 |
| 18aFLR | 23,355,062 | leaf | female | RNA | SAMN20803195 | SRR15881662 |
| 19eFLR | 23,713,364 | leaf | female | RNA | SAMN20803196 | SRR15881651 |
| 1dFLR | 22,062,003 | leaf | female | RNA | SAMN20803256 | SRR15881674 |
| 20bFLR | 24,065,165 | leaf | female | RNA | SAMN20803197 | SRR15881868 |
| 22eFLR | 29,235,803 | leaf | female | RNA | SAMN20803198 | SRR15881857 |
| 23fFLR | 21,076,535 | leaf | female | RNA | SAMN20803199 | SRR15881846 |
| 24aFLR | 28,749,957 | leaf | female | RNA | SAMN20803200 | SRR15881835 |
| 25aFLR | 27,871,738 | leaf | female | RNA | SAMN20803201 | SRR15881824 |
| 27cFLR | 22,161,951 | leaf | female | RNA | SAMN20803202 | SRR15881813 |
| 28cFLR | 23,213,856 | leaf | female | RNA | SAMN20803203 | SRR15881802 |
| 29cFLR | 21,196,131 | leaf | female | RNA | SAMN20803204 | SRR15881791 |
| 2fFLR | 16,923,579 | leaf | female | RNA | SAMN20803259 | SRR15881670 |
| 30bFLR | 21,417,585 | leaf | female | RNA | SAMN20803205 | SRR15881780 |
| 31dFLR | 19,806,817 | leaf | female | RNA | SAMN20803206 | SRR15881769 |
| 32fFLR | 28,376,256 | leaf | female | RNA | SAMN20803207 | SRR15881757 |
| 33fFLR | 26,787,156 | leaf | female | RNA | SAMN20803208 | SRR15881746 |

| 34fFLR | 19,936,395 | leaf | female | RNA | SAMN20803209 | SRR15881735 |
|--------|------------|------|--------|-----|--------------|-------------|
| 35bFLR | 25,090,395 | leaf | female | RNA | SAMN20803210 | SRR15881724 |
| 36dFLR | 35,664,150 | leaf | female | RNA | SAMN20803211 | SRR15881723 |
| 38dFLR | 28,092,357 | leaf | female | RNA | SAMN20803212 | SRR15881722 |
| 39dFLR | 23,414,246 | leaf | female | RNA | SAMN20803213 | SRR15881721 |
| 3bFLR | 24,338,736 | leaf | female | RNA | SAMN20803260 | SRR15881669 |
| 40aFLR | 38,984,947 | leaf | female | RNA | SAMN20803214 | SRR15881720 |
| 41cFLR | 23,437,737 | leaf | female | RNA | SAMN20803215 | SRR15881719 |
| 42dFLR | 30,700,908 | leaf | female | RNA | SAMN20803216 | SRR15881718 |
| 43aFLR | 22,120,081 | leaf | female | RNA | SAMN20803217 | SRR15881716 |
| 44fFLR | 24,371,277 | leaf | female | RNA | SAMN20803218 | SRR15881715 |
| 45aFLR | 15,439,345 | leaf | female | RNA | SAMN20803219 | SRR15881714 |
| 46aFLR | 21,179,572 | leaf | female | RNA | SAMN20803220 | SRR15881713 |
| 47fFLR | 26,757,888 | leaf | female | RNA | SAMN20803221 | SRR15881712 |
| 48cFLR | 24,020,971 | leaf | female | RNA | SAMN20803222 | SRR15881711 |
| 49bFLR | 25,110,204 | leaf | female | RNA | SAMN20803223 | SRR15881710 |
| 4cFLR | 21,497,866 | leaf | female | RNA | SAMN20803263 | SRR15881666 |
| 50aFLR | 20,252,697 | leaf | female | RNA | SAMN20803224 | SRR15881709 |
| 51fFLR | 24,709,224 | leaf | female | RNA | SAMN20803225 | SRR15881708 |
| 52cFLR | 20,680,749 | leaf | female | RNA | SAMN20803226 | SRR15881707 |
| 53cFLR | 24,590,164 | leaf | female | RNA | SAMN20803227 | SRR15881705 |
| 54aFLR | 22,070,920 | leaf | female | RNA | SAMN20803228 | SRR15881704 |
| 55cFLR | 21,520,506 | leaf | female | RNA | SAMN20803229 | SRR15881703 |
| 56aFLR | 19,082,140 | leaf | female | RNA | SAMN20803230 | SRR15881702 |
| 57cFLR | 25,707,837 | leaf | female | RNA | SAMN20803231 | SRR15881701 |
| 58bFLR | 21,853,794 | leaf | female | RNA | SAMN20803232 | SRR15881700 |
| 59aFLR | 27,249,071 | leaf | female | RNA | SAMN20803233 | SRR15881699 |
| 5dFLR | 20,563,422 | leaf | female | RNA | SAMN20803265 | SRR15881664 |
| 60cFLR | 34,953,616 | leaf | female | RNA | SAMN20803234 | SRR15881698 |
| 61fFLR | 30,145,219 | leaf | female | RNA | SAMN20803235 | SRR15881697 |
| 62dFLR | 21,552,687 | leaf | female | RNA | SAMN20803236 | SRR15881696 |
| 63fFLR | 23,486,608 | leaf | female | RNA | SAMN20803237 | SRR15881694 |
| 64aFLR | 26,336,128 | leaf | female | RNA | SAMN20803238 | SRR15881693 |
| 65dFLR | 25,427,194 | leaf | female | RNA | SAMN20803239 | SRR15881692 |
| 66aFLR | 25,823,184 | leaf | female | RNA | SAMN20803240 | SRR15881691 |
| 67eFLR | 23,625,471 | leaf | female | RNA | SAMN20803241 | SRR15881690 |
| 68cFLR | 38,995,391 | leaf | female | RNA | SAMN20803242 | SRR15881689 |
| 69aFLR | 22,340,663 | leaf | female | RNA | SAMN20803243 | SRR15881688 |
| 6aFLR | 20,027,447 | leaf | female | RNA | SAMN20803266 | SRR15881663 |
| 70fFLR | 22,136,670 | leaf | female | RNA | SAMN20803244 | SRR15881687 |
| 71aFLR | 25,277,467 | leaf | female | RNA | SAMN20803245 | SRR15881686 |
| 72dFLR | 24,064,655 | leaf | female | RNA | SAMN20803246 | SRR15881685 |
| 73dFLR | 20,270,778 | leaf | female | RNA | SAMN20803247 | SRR15881683 |
| 74eFLR | 22,579,935 | leaf | female | RNA | SAMN20803248 | SRR15881682 |
| 75dFLR | 27,424,320 | leaf | female | RNA | SAMN20803249 | SRR15881681 |
| 76dFLR | 22,422,304 | leaf | female | RNA | SAMN20803250 | SRR15881680 |

| 77fFLR | 13,013,854 | leaf | female | RNA | SAMN20803251 | SRR15881679 |
| 78aFLR | 22,098,803 | leaf | female | RNA | SAMN20803252 | SRR15881678 |
| 79aFLR | 20,942,525 | leaf | female | RNA | SAMN20803253 | SRR15881677 |
| 7cFLR | 20,784,675 | leaf | female | RNA | SAMN20803269 | SRR15881659 |
| 80bFLR | 20,968,381 | leaf | female | RNA | SAMN20803254 | SRR15881676 |
| 81cFLR | 20,576,875 | leaf | female | RNA | SAMN20803255 | SRR15881675 |
| 8aFLR | 23,068,794 | leaf | female | RNA | SAMN20803270 | SRR15881658 |
| 9cFLR | 29,862,383 | leaf | female | RNA | SAMN20803272 | SRR15881656 |

Table A2. *Rumex hastatulus* male leaf and pollen RNA sequencing data.

| Library ID | NCBI SRA Run Accession ID | NCBI SRA Accession ID | Biosample accession | Tissue | Sex | Read Number |
|---|---|---|---|---|---|---|
| 30fMLR | SRR15881842 | SRP336884 | SAMN20803310 | leaf | male | 18,618,763 |
| 56eMLR | SRR15881785 | SRP336884 | SAMN20803362 | leaf | male | 19,223,242 |
| 32aMLR | SRR15881838 | SRP336884 | SAMN20803314 | leaf | male | 19,224,076 |
| 40bMLR | SRR15881820 | SRP336884 | SAMN20803330 | leaf | male | 19,260,622 |
| 77bMLR | SRR15881743 | SRP336884 | SAMN20803399 | leaf | male | 19,374,457 |
| 33aMLR | SRR15881836 | SRP336884 | SAMN20803316 | leaf | male | 19,505,134 |
| 45dMLR | SRR15881809 | SRP336884 | SAMN20803340 | leaf | male | 19,581,862 |
| 27eMLR | SRR15881849 | SRP336884 | SAMN20803304 | leaf | male | 19,678,341 |
| 23aMPR | SRR15881855 | SRP336884 | SAMN20803298 | pollen | male | 19,900,577 |
| 52eMPR | SRR15881793 | SRP336884 | SAMN20803355 | pollen | male | 20,068,166 |
| 47dMPR | SRR15881804 | SRP336884 | SAMN20803345 | pollen | male | 20,137,178 |
| 51dMPR | SRR15881795 | SRP336884 | SAMN20803353 | pollen | male | 20,354,254 |
| 63aMLR | SRR15881770 | SRP336884 | SAMN20803376 | leaf | male | 20,366,263 |
| 46eMLR | SRR15881807 | SRP336884 | SAMN20803342 | leaf | male | 20,405,406 |
| 74bMPR | SRR15881749 | SRP336884 | SAMN20803394 | pollen | male | 20,469,683 |
| 48aMLR | SRR15881803 | SRP336884 | SAMN20803346 | leaf | male | 20,499,705 |
| 20aMLR | SRR15881861 | SRP336884 | SAMN20803293 | leaf | male | 20,648,435 |
| 18cMLR | SRR15881865 | SRP336884 | SAMN20803289 | leaf | male | 20,719,137 |
| 59cMLR | SRR15881778 | SRP336884 | SAMN20803368 | leaf | male | 20,787,752 |
| 16bMPR | SRR15881641 | SRP336884 | SAMN20803286 | pollen | male | 20,816,666 |
| 17bMPR | SRR15881866 | SRP336884 | SAMN20803288 | pollen | male | 20,905,752 |
| 57aMLR | SRR15881783 | SRP336884 | SAMN20803364 | leaf | male | 20,949,447 |
| 41eMLR | SRR15881818 | SRP336884 | SAMN20803332 | leaf | male | 21,028,048 |
| 2bMLR | SRR15881671 | SRP336884 | SAMN20803258 | leaf | male | 21,029,703 |
| 58fMLR | SRR15881781 | SRP336884 | SAMN20803366 | leaf | male | 21,088,467 |
| 15dMPR | SRR15881643 | SRP336884 | SAMN20803284 | pollen | male | 21,208,157 |
| 75fMLR | SRR15881748 | SRP336884 | SAMN20803395 | leaf | male | 21,248,272 |
| 66dMPR | SRR15881764 | SRP336884 | SAMN20803381 | pollen | male | 21,287,121 |
| 47dMLR | SRR15881805 | SRP336884 | SAMN20803344 | leaf | male | 21,305,232 |
| 36cMLR | SRR15881829 | SRP336884 | SAMN20803322 | leaf | male | 21,314,930 |
| 13aMPR | SRR15881647 | SRP336884 | SAMN20803280 | pollen | male | 21,393,522 |

| 2bMPR | SRR15881731 | SRP336884 | SAMN20803410 | pollen | male | 21,431,603 |
|---|---|---|---|---|---|---|
| 66dMLR | SRR15881765 | SRP336884 | SAMN20803380 | leaf | male | 21,449,917 |
| 31eMLR | SRR15881840 | SRP336884 | SAMN20803312 | leaf | male | 21,491,348 |
| 14eMPR | SRR15881645 | SRP336884 | SAMN20803282 | pollen | male | 21,495,818 |
| 49fMLR | SRR15881800 | SRP336884 | SAMN20803348 | leaf | male | 21,648,198 |
| 11aMPR | SRR15881652 | SRP336884 | SAMN20803276 | pollen | male | 21,672,646 |
| 27eMPR | SRR15881848 | SRP336884 | SAMN20803305 | pollen | male | 21,730,633 |
| 9bMPR | SRR15881725 | SRP336884 | SAMN20803416 | pollen | male | 21,736,155 |
| 38bMLR | SRR15881825 | SRP336884 | SAMN20803326 | leaf | male | 21,737,917 |
| 67fMLR | SRR15881763 | SRP336884 | SAMN20803382 | leaf | male | 21,798,565 |
| 54bMLR | SRR15881789 | SRP336884 | SAMN20803358 | leaf | male | 21,801,814 |
| 48aMPR | SRR15881801 | SRP336884 | SAMN20803347 | pollen | male | 21,879,161 |
| 59cMPR | SRR15881777 | SRP336884 | SAMN20803369 | pollen | male | 21,891,858 |
| 68eMPR | SRR15881760 | SRP336884 | SAMN20803385 | pollen | male | 22,014,750 |
| 33aMPR | SRR15881834 | SRP336884 | SAMN20803317 | pollen | male | 22,232,033 |
| 37fMPR | SRR15881826 | SRP336884 | SAMN20803325 | pollen | male | 22,245,344 |
| 24fMPR | SRR15881853 | SRP336884 | SAMN20803300 | pollen | male | 22,250,004 |
| 80aMLR | SRR15881737 | SRP336884 | SAMN20803405 | leaf | male | 22,285,796 |
| 56eMPR | SRR15881784 | SRP336884 | SAMN20803363 | pollen | male | 22,350,802 |
| 70bMLR | SRR15881755 | SRP336884 | SAMN20803388 | leaf | male | 22,381,085 |
| 61cMPR | SRR15881773 | SRP336884 | SAMN20803373 | pollen | male | 22,478,607 |
| 12fMPR | SRR15881649 | SRP336884 | SAMN20803278 | pollen | male | 22,558,410 |
| 39fMLR | SRR15881822 | SRP336884 | SAMN20803328 | leaf | male | 22,613,466 |
| 1fMPR | SRR15881732 | SRP336884 | SAMN20803409 | pollen | male | 22,618,797 |
| 67fMPR | SRR15881762 | SRP336884 | SAMN20803383 | pollen | male | 22,624,892 |
| 49fMPR | SRR15881799 | SRP336884 | SAMN20803349 | pollen | male | 22,646,902 |
| 29eMPR | SRR15881843 | SRP336884 | SAMN20803309 | pollen | male | 22,676,163 |
| 77bMPR | SRR15881742 | SRP336884 | SAMN20803400 | pollen | male | 22,747,342 |
| 53bMLR | SRR15881792 | SRP336884 | SAMN20803356 | leaf | male | 22,752,949 |
| 43bMPR | SRR15881812 | SRP336884 | SAMN20803337 | pollen | male | 22,846,342 |
| 55dMPR | SRR15881786 | SRP336884 | SAMN20803361 | pollen | male | 22,872,080 |
| 53bMPR | SRR15881790 | SRP336884 | SAMN20803357 | pollen | male | 22,889,516 |
| 7bMPR | SRR15881726 | SRP336884 | SAMN20803415 | pollen | male | 22,968,783 |
| 4aMPR | SRR15881729 | SRP336884 | SAMN20803412 | pollen | male | 23,008,155 |
| 19aMPR | SRR15881862 | SRP336884 | SAMN20803292 | pollen | male | 23,052,887 |
| 57aMPR | SRR15881782 | SRP336884 | SAMN20803365 | pollen | male | 23,080,691 |
| 71eMPR | SRR15881752 | SRP336884 | SAMN20803391 | pollen | male | 23,140,418 |
| 20aMPR | SRR15881860 | SRP336884 | SAMN20803294 | pollen | male | 23,176,473 |
| 23aMLR | SRR15881856 | SRP336884 | SAMN20803297 | leaf | male | 23,231,005 |
| 44bMPR | SRR15881810 | SRP336884 | SAMN20803339 | pollen | male | 23,253,134 |
| 35aMLR | SRR15881831 | SRP336884 | SAMN20803320 | leaf | male | 23,270,364 |
| 44bMLR | SRR15881811 | SRP336884 | SAMN20803338 | leaf | male | 23,274,185 |
| 69dMPR | SRR15881756 | SRP336884 | SAMN20803387 | pollen | male | 23,323,530 |
| 78eMPR | SRR15881740 | SRP336884 | SAMN20803402 | pollen | male | 23,337,145 |
| 79bMPR | SRR15881738 | SRP336884 | SAMN20803404 | pollen | male | 23,352,267 |
| 26fMLR | SRR15881850 | SRP336884 | SAMN20803303 | leaf | male | 23,473,409 |

| 81dMLR | SRR15881734 | SRP336884 | SAMN20803407 | leaf | male | 23,574,338 |
|--------|-------------|-----------|--------------|------|------|------------|
| 60fMLR | SRR15881776 | SRP336884 | SAMN20803370 | leaf | male | 23,622,003 |
| 42cMLR | SRR15881816 | SRP336884 | SAMN20803334 | leaf | male | 23,623,400 |
| 22dMPR | SRR15881858 | SRP336884 | SAMN20803296 | pollen | male | 23,649,737 |
| 30fMPR | SRR15881841 | SRP336884 | SAMN20803311 | pollen | male | 23,685,258 |
| 62bMPR | SRR15881771 | SRP336884 | SAMN20803375 | pollen | male | 23,878,383 |
| 34cMLR | SRR15881833 | SRP336884 | SAMN20803318 | leaf | male | 23,883,801 |
| 18cMPR | SRR15881864 | SRP336884 | SAMN20803290 | pollen | male | 23,886,673 |
| 68eMLR | SRR15881761 | SRP336884 | SAMN20803384 | leaf | male | 23,891,242 |
| 9bMLR | SRR15881657 | SRP336884 | SAMN20803271 | leaf | male | 23,903,205 |
| 11aMLR | SRR15881653 | SRP336884 | SAMN20803275 | leaf | male | 23,950,560 |
| 60fMPR | SRR15881775 | SRP336884 | SAMN20803371 | pollen | male | 23,977,717 |
| 73aMLR | SRR15881751 | SRP336884 | SAMN20803392 | leaf | male | 24,014,241 |
| 55dMLR | SRR15881787 | SRP336884 | SAMN20803360 | leaf | male | 24,072,233 |
| 10aMPR | SRR15881654 | SRP336884 | SAMN20803274 | pollen | male | 24,080,539 |
| 6dMPR | SRR15881727 | SRP336884 | SAMN20803414 | pollen | male | 24,110,678 |
| 80aMPR | SRR15881736 | SRP336884 | SAMN20803406 | pollen | male | 24,135,007 |
| 35aMPR | SRR15881830 | SRP336884 | SAMN20803321 | pollen | male | 24,210,257 |
| 81dMPR | SRR15881733 | SRP336884 | SAMN20803408 | pollen | male | 24,217,871 |
| 29eMLR | SRR15881844 | SRP336884 | SAMN20803308 | leaf | male | 24,228,138 |
| 19aMLR | SRR15881863 | SRP336884 | SAMN20803291 | leaf | male | 24,308,115 |
| 62bMLR | SRR15881772 | SRP336884 | SAMN20803374 | leaf | male | 24,341,141 |
| 58fMPR | SRR15881779 | SRP336884 | SAMN20803367 | pollen | male | 24,388,519 |
| 3dMPR | SRR15881730 | SRP336884 | SAMN20803411 | pollen | male | 24,454,526 |
| 31eMPR | SRR15881839 | SRP336884 | SAMN20803313 | pollen | male | 24,467,269 |
| 46eMPR | SRR15881806 | SRP336884 | SAMN20803343 | pollen | male | 24,476,988 |
| 41eMPR | SRR15881817 | SRP336884 | SAMN20803333 | pollen | male | 24,499,761 |
| 25dMPR | SRR15881851 | SRP336884 | SAMN20803302 | pollen | male | 24,576,626 |
| 5aMLR | SRR15881665 | SRP336884 | SAMN20803264 | leaf | male | 24,647,049 |
| 76bMLR | SRR15881745 | SRP336884 | SAMN20803397 | leaf | male | 24,796,587 |
| 50bMPR | SRR15881797 | SRP336884 | SAMN20803351 | pollen | male | 24,806,804 |
| 75fMPR | SRR15881747 | SRP336884 | SAMN20803396 | pollen | male | 24,967,559 |
| 17bMLR | SRR15881867 | SRP336884 | SAMN20803287 | leaf | male | 25,097,622 |
| 36cMPR | SRR15881828 | SRP336884 | SAMN20803323 | pollen | male | 25,117,025 |
| 65cMPR | SRR15881766 | SRP336884 | SAMN20803379 | pollen | male | 25,212,386 |
| 40bMPR | SRR15881819 | SRP336884 | SAMN20803331 | pollen | male | 25,218,392 |
| 14eMLR | SRR15881646 | SRP336884 | SAMN20803281 | leaf | male | 25,242,708 |
| 3dMLR | SRR15881668 | SRP336884 | SAMN20803261 | leaf | male | 25,372,184 |
| 34cMPR | SRR15881832 | SRP336884 | SAMN20803319 | pollen | male | 25,379,543 |
| 52eMLR | SRR15881794 | SRP336884 | SAMN20803354 | leaf | male | 25,400,339 |
| 1fMLR | SRR15881672 | SRP336884 | SAMN20803257 | leaf | male | 25,687,702 |
| 22dMLR | SRR15881859 | SRP336884 | SAMN20803295 | leaf | male | 25,806,398 |
| 65cMLR | SRR15881767 | SRP336884 | SAMN20803378 | leaf | male | 25,830,897 |
| 16bMLR | SRR15881642 | SRP336884 | SAMN20803285 | leaf | male | 26,035,066 |
| 5aMPR | SRR15881728 | SRP336884 | SAMN20803413 | pollen | male | 26,042,949 |
| 28dMPR | SRR15881845 | SRP336884 | SAMN20803307 | pollen | male | 26,049,346 |

| 63aMPR | SRR15881768 | SRP336884 | SAMN20803377 | pollen | male | 26,062,542 |
|--------|-------------|-----------|--------------|--------|------|------------|
| 28dMLR | SRR15881847 | SRP336884 | SAMN20803306 | leaf | male | 26,162,341 |
| 7bMLR | SRR15881660 | SRP336884 | SAMN20803268 | leaf | male | 26,243,453 |
| 45dMPR | SRR15881808 | SRP336884 | SAMN20803341 | pollen | male | 26,299,368 |
| 32aMPR | SRR15881837 | SRP336884 | SAMN20803315 | pollen | male | 26,315,190 |
| 78eMLR | SRR15881741 | SRP336884 | SAMN20803401 | leaf | male | 27,145,132 |
| 6dMLR | SRR15881661 | SRP336884 | SAMN20803267 | leaf | male | 27,358,838 |
| 37fMLR | SRR15881827 | SRP336884 | SAMN20803324 | leaf | male | 27,393,508 |
| 12fMLR | SRR15881650 | SRP336884 | SAMN20803277 | leaf | male | 27,995,721 |
| 25dMLR | SRR15881852 | SRP336884 | SAMN20803301 | leaf | male | 28,050,370 |
| 24fMLR | SRR15881854 | SRP336884 | SAMN20803299 | leaf | male | 28,270,981 |
| 4aMLR | SRR15881667 | SRP336884 | SAMN20803262 | leaf | male | 28,281,221 |
| 76bMPR | SRR15881744 | SRP336884 | SAMN20803398 | pollen | male | 28,684,678 |
| 79bMLR | SRR15881739 | SRP336884 | SAMN20803403 | leaf | male | 28,729,279 |
| 15dMLR | SRR15881644 | SRP336884 | SAMN20803283 | leaf | male | 28,967,120 |
| 51dMLR | SRR15881796 | SRP336884 | SAMN20803352 | leaf | male | 29,400,094 |
| 10aMLR | SRR15881655 | SRP336884 | SAMN20803273 | leaf | male | 30,132,989 |
| 39fMPR | SRR15881821 | SRP336884 | SAMN20803329 | pollen | male | 30,824,690 |
| 42cMPR | SRR15881815 | SRP336884 | SAMN20803335 | pollen | male | 31,080,825 |
| 69dMLR | SRR15881759 | SRP336884 | SAMN20803386 | leaf | male | 32,288,528 |
| 54bMPR | SRR15881788 | SRP336884 | SAMN20803359 | pollen | male | 32,291,999 |
| 38bMPR | SRR15881823 | SRP336884 | SAMN20803327 | pollen | male | 32,438,259 |
| 61cMLR | SRR15881774 | SRP336884 | SAMN20803372 | leaf | male | 32,570,831 |
| 13aMLR | SRR15881648 | SRP336884 | SAMN20803279 | leaf | male | 34,475,828 |
| 50bMLR | SRR15881798 | SRP336884 | SAMN20803350 | leaf | male | 36,328,300 |
| 71eMLR | SRR15881753 | SRP336884 | SAMN20803390 | leaf | male | 36,606,732 |
| 74bMLR | SRR15881750 | SRP336884 | SAMN20803393 | leaf | male | 36,617,245 |
| 43bMLR | SRR15881814 | SRP336884 | SAMN20803336 | leaf | male | 51,113,749 |
| 70bMPR | SRR15881754 | SRP336884 | SAMN20803389 | pollen | male | 58,151,564 |

Table A3. Number of DE genes in *Rumex hastatulus*. *P*-values were adjusted with an FDR of 0.1. Genes with a mean raw read count < 5 across all samples were removed. FC: fold change in expression level.

| | | leaf-biased | | | pollen-biased | |
|---|---|---|---|---|---|---|
| | p<0.05 | p<0.05, FC >2 | p<0.05, FC >4 | p<0.05 | p<0.05, FC >2 | p<0.05, FC >4 |
| **Autosome** | 9131 | 7695 | 4887 | 6127 | 5135 | 3990 |
| **PAR** | 1137 | 981 | 641 | 722 | 610 | 469 |
| **X** | 689 | 543 | 351 | 497 | 422 | 331 |
| **Y** | 874 | 728 | 418 | 645 | 535 | 415 |
| **Total** | 11831 | 9947 | 6297 | 7991 | 6702 | 5205 |
| | | female-biased | | | male-biased | |
| | p<0.05 | p<0.05, FC >2 | p<0.05, FC >4 | p<0.05 | p<0.05, FC >2 | p<0.05, FC >4 |
| **Autosome** | 604 | 10 | 1 | 1872 | 50 | 3 |
| **PAR** | 41 | 2 | 2 | 174 | 11 | 2 |
| **X** | 1628 | 378 | 2 | 5 | 4 | 3 |
| **Y** | 42 | 15 | 0 | 1270 | 1026 | 747 |
| **Total** | 2315 | 405 | 5 | 3321 | 1091 | 755 |

Table A4. GO enrichment of eGenes in male leaf, female leaf, and pollen of *Rumex hastatulus*.

**Male leaf**

| GO.ID | Term | Annotated | Significant | Expected | P-value |
|---|---|---|---|---|---|
| GO:0000413 | protein peptidyl-prolyl isomerization | 19 | 7 | 2.55 | 0.0089 |
| GO:0006812 | cation transport | 100 | 16 | 13.45 | 0.0089 |
| GO:0006520 | cellular amino acid metabolic process | 107 | 19 | 14.39 | 0.015 |
| GO:0016579 | protein deubiquitination | 21 | 7 | 2.82 | 0.0162 |
| GO:0071705 | nitrogen compound transport | 105 | 17 | 14.12 | 0.018 |
| GO:0022607 | cellular component assembly | 55 | 10 | 7.4 | 0.0214 |
| GO:0019752 | carboxylic acid metabolic process | 180 | 32 | 24.2 | 0.0472 |

**Female leaf**

| GO.ID | Term | Annotated | Significant | Expected | P-value |
|---|---|---|---|---|---|
| GO:0033036 | macromolecule localization | 110 | 30 | 22.37 | 0.012 |
| GO:0006281 | DNA repair | 43 | 15 | 8.74 | 0.018 |
| GO:0019538 | protein metabolic process | 1162 | 235 | 236.29 | 0.029 |
| GO:0006913 | nucleocytoplasmic transport | 10 | 5 | 2.03 | 0.035 |
| GO:0006952 | defense response | 24 | 9 | 4.88 | 0.039 |
| GO:0006796 | phosphate-containing compound metabolic process | 761 | 176 | 154.75 | 0.04 |
| GO:0071705 | nitrogen compound transport | 105 | 27 | 21.35 | 0.041 |
| GO:0009312 | oligosaccharide biosynthetic process | 19 | 8 | 3.86 | 0.041 |
| GO:0032259 | methylation | 19 | 4 | 3.86 | 0.041 |
| GO:0051252 | regulation of RNA metabolic process | 361 | 55 | 73.41 | 0.044 |
| GO:0009152 | purine ribonucleotide biosynthetic process | 14 | 6 | 2.85 | 0.047 |

**Pollen**

| GO.ID | Term | Annotated | Significant | Expected | P-value |
|---|---|---|---|---|---|
| GO:0046700 | heterocycle catabolic process | 13 | 5 | 1.04 | 0.0024 |
| GO:0019439 | aromatic compound catabolic process | 14 | 5 | 1.12 | 0.0034 |
| GO:0044270 | cellular nitrogen compound catabolic process | 14 | 5 | 1.12 | 0.0034 |
| GO:1901361 | organic cyclic compound catabolic process | 14 | 5 | 1.12 | 0.0034 |
| GO:0006643 | membrane lipid metabolic process | 10 | 4 | 0.8 | 0.0057 |
| GO:0051252 | regulation of RNA metabolic process | 314 | 20 | 25.09 | 0.0067 |
| GO:0016071 | mRNA metabolic process | 26 | 5 | 2.08 | 0.0084 |
| GO:0010629 | negative regulation of gene expression | 11 | 4 | 0.88 | 0.0084 |
| GO:0048518 | positive regulation of biological process | 12 | 4 | 0.96 | 0.0118 |
| GO:0006352 | DNA-templated transcription initiation | 19 | 5 | 1.52 | 0.0144 |
| GO:0008654 | phospholipid biosynthetic process | 20 | 5 | 1.6 | 0.0179 |
| GO:0010556 | regulation of macromolecule biosynthetic process | 315 | 20 | 25.17 | 0.0191 |
| GO:0031326 | regulation of cellular biosynthetic process | 315 | 20 | 25.17 | 0.0191 |
| GO:0044265 | cellular macromolecule catabolic process | 42 | 6 | 3.36 | 0.0208 |
| GO:0016579 | protein deubiquitination | 21 | 5 | 1.68 | 0.022 |
| GO:0065008 | regulation of biological quality | 22 | 5 | 1.76 | 0.0267 |
| GO:0071805 | potassium ion transmembrane transport | 15 | 4 | 1.2 | 0.0269 |

| GO:0050896 | response to stimulus | 226 | 24 | 18.06 | 0.0331 |
|---|---|---|---|---|---|
| GO:0033036 | macromolecule localization | 109 | 13 | 8.71 | 0.0389 |
| GO:0042440 | pigment metabolic process | 10 | 3 | 0.8 | 0.0397 |

Table A5. GO enrichment of eGenes with discordant eQTLs between life-stages or sexes in *Rumex hastatulus*.

| Discordant between sexes | | | | | |
|---|---|---|---|---|---|
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P-value** |
| GO:0010629 | negative regulation of gene expression | 12 | 4 | 0.68 | 0.0035 |
| GO:0051252 | regulation of RNA metabolic process | 396 | 13 | 22.43 | 0.0036 |
| GO:0006139 | nucleobase-containing compound metabolic process | 773 | 40 | 43.79 | 0.0059 |
| GO:0044265 | cellular macromolecule catabolic process | 43 | 5 | 2.44 | 0.0081 |
| GO:0010556 | regulation of macromolecule biosynthetic process | 397 | 13 | 22.49 | 0.0103 |
| GO:0031326 | regulation of cellular biosynthetic process | 397 | 13 | 22.49 | 0.0103 |
| GO:0033036 | macromolecule localization | 112 | 10 | 6.34 | 0.0207 |
| GO:0051246 | regulation of protein metabolic process | 11 | 3 | 0.62 | 0.0211 |
| GO:0016311 | dephosphorylation | 20 | 4 | 1.13 | 0.0237 |
| GO:0050790 | regulation of catalytic activity | 12 | 3 | 0.68 | 0.027 |
| GO:0044282 | small molecule catabolic process | 13 | 3 | 0.74 | 0.0336 |
| GO:0046700 | heterocycle catabolic process | 13 | 3 | 0.74 | 0.0336 |
| GO:0019439 | aromatic compound catabolic process | 14 | 3 | 0.79 | 0.041 |
| GO:0009152 | purine ribonucleotide biosynthetic process | 14 | 3 | 0.79 | 0.041 |
| GO:0044270 | cellular nitrogen compound catabolic process | 14 | 3 | 0.79 | 0.041 |
| GO:1901361 | organic cyclic compound catabolic process | 14 | 3 | 0.79 | 0.041 |
| GO:0048518 | positive regulation of biological process | 14 | 3 | 0.79 | 0.041 |
| **Discordant between life-stages** | | | | | |
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P-value** |
| GO:0005984 | disaccharide metabolic process | 26 | 3 | 0.27 | 0.0028 |
| GO:0006355 | regulation of DNA-templated transcription | 369 | 7 | 3.87 | 0.085 |

Table A6. GO enrichment of eGenes with discordant eQTLs for expression differences between life-stages of Rumex hastatulus.

| GO.ID | Term | Annotated | Significant | Expected | P-value |
|---|---|---|---|---|---|
| GO:0034641 | cellular nitrogen compound metabolic process | 1058 | 66 | 70.63 | 0.00033 |
| GO:0051252 | regulation of RNA metabolic process | 396 | 18 | 26.44 | 0.00485 |
| GO:0000413 | protein peptidyl-prolyl isomerization | 19 | 5 | 1.27 | 0.00682 |
| GO:0046700 | heterocycle catabolic process | 13 | 4 | 0.87 | 0.00857 |
| GO:0009308 | amine metabolic process | 13 | 4 | 0.87 | 0.00857 |
| GO:0019439 | aromatic compound catabolic process | 14 | 4 | 0.93 | 0.01138 |
| GO:0044270 | cellular nitrogen compound catabolic process | 14 | 4 | 0.93 | 0.01138 |
| GO:1901361 | organic cyclic compound catabolic process | 14 | 4 | 0.93 | 0.01138 |
| GO:0046483 | heterocycle metabolic process | 816 | 52 | 54.47 | 0.01367 |
| GO:0010556 | regulation of macromolecule biosynthetic process | 397 | 18 | 26.5 | 0.01389 |
| GO:0031326 | regulation of cellular biosynthetic process | 397 | 18 | 26.5 | 0.01389 |
| GO:0006749 | glutathione metabolic process | 31 | 6 | 2.07 | 0.0149 |
| GO:1901565 | organonitrogen compound catabolic process | 62 | 7 | 4.14 | 0.01864 |
| GO:0044248 | cellular catabolic process | 63 | 10 | 4.21 | 0.0313 |
| GO:0010629 | negative regulation of gene expression | 12 | 3 | 0.8 | 0.04124 |
| GO:0009072 | aromatic amino acid family metabolic process | 12 | 3 | 0.8 | 0.04124 |
| GO:0006364 | rRNA processing | 21 | 4 | 1.4 | 0.04717 |

# Chapter 2 Supplementary Figures
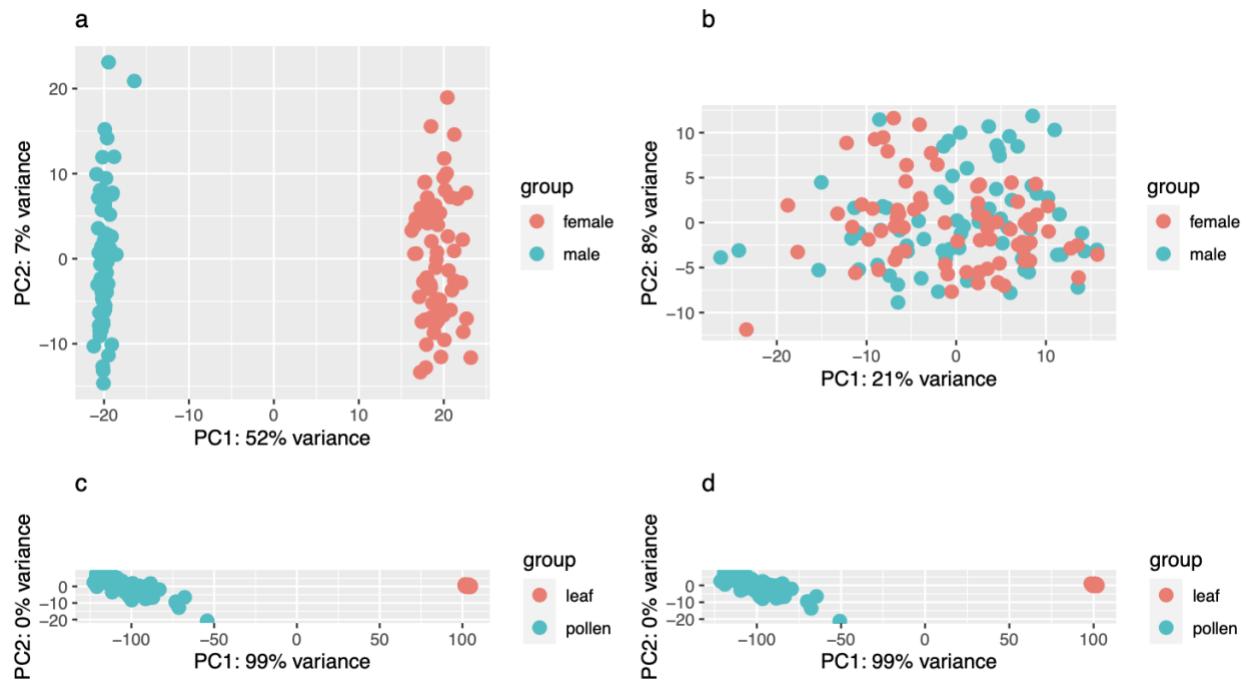


Figure A1. PCA of leaf and pollen RNA samples of *Rumex hastatulus*. a-b: male vs. female leaf. c-d: pollen vs. male leaf. a,c: whole genome including X- and Y-specific regions, b,d: only autosomes and PAR.
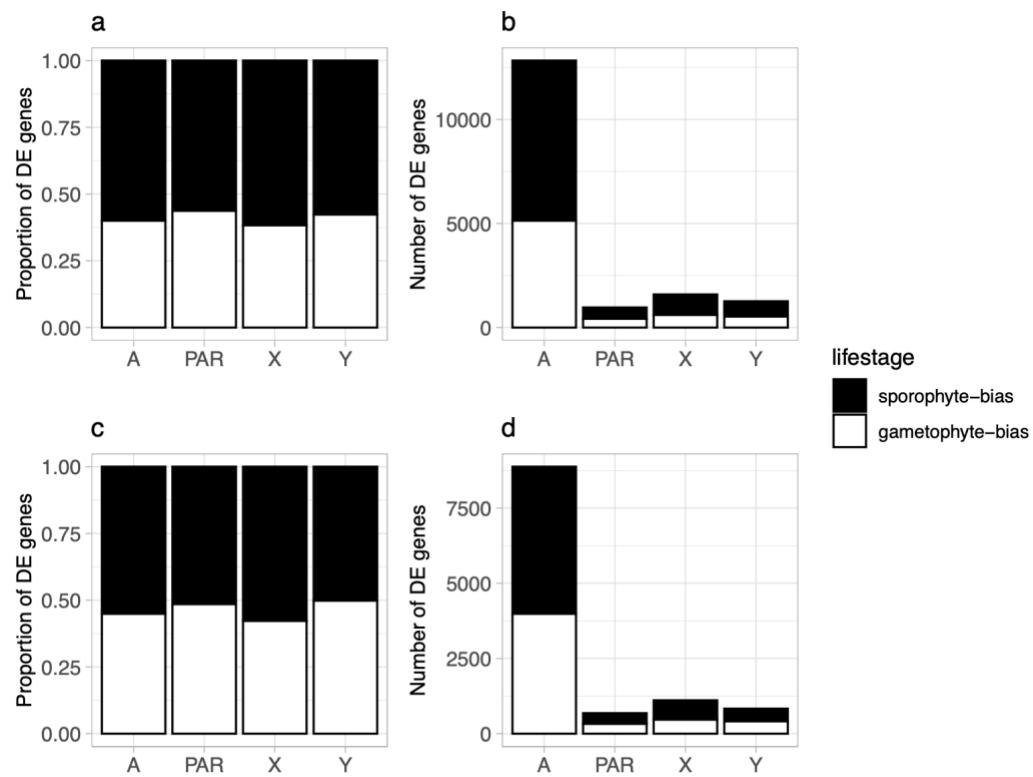
Figure A2. Enrichment of gametophyte-biased genes on the sex chromosomes in *Rumex hastatulus*. Cutoff for DE genes: adjusted *p* < 0.05 and fold change > 2 (a, b), fold change > 4 (c, d).
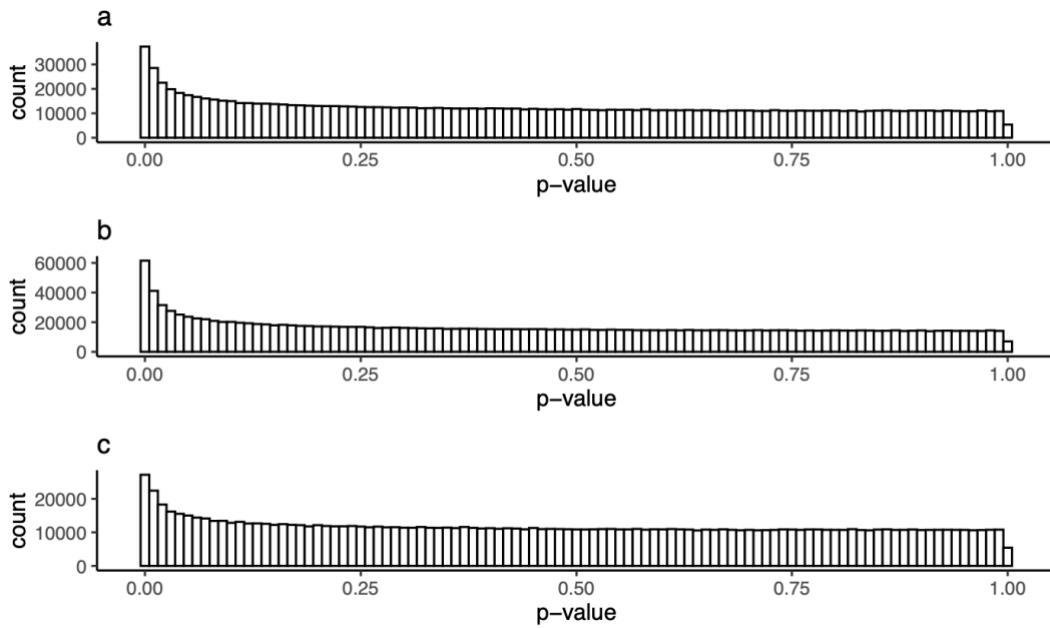
Figure A3. Distribution of nominal *p*-values for all nearby SNPs tested in male leaf (a), female leaf (b), pollen (c) of *Rumex hastatulus*. Bin size = 0.01.
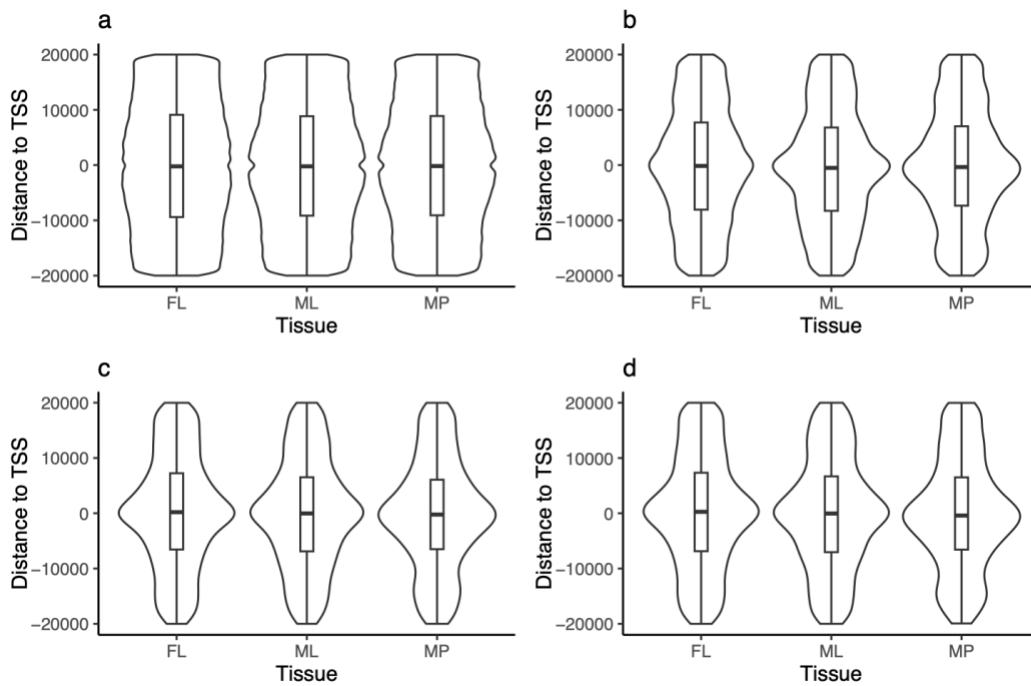


Figure A4. Distance between eQTL and the transcription start site of eQTLs male leaf (ML), female leaf (FL), pollen (MP) of *Rumex hastatulus*. a. All nearby SNPs, b. all eQTLs, c. the most significant eQTL for each eGene, d. a randomly selected significant eQTL for each eGene.
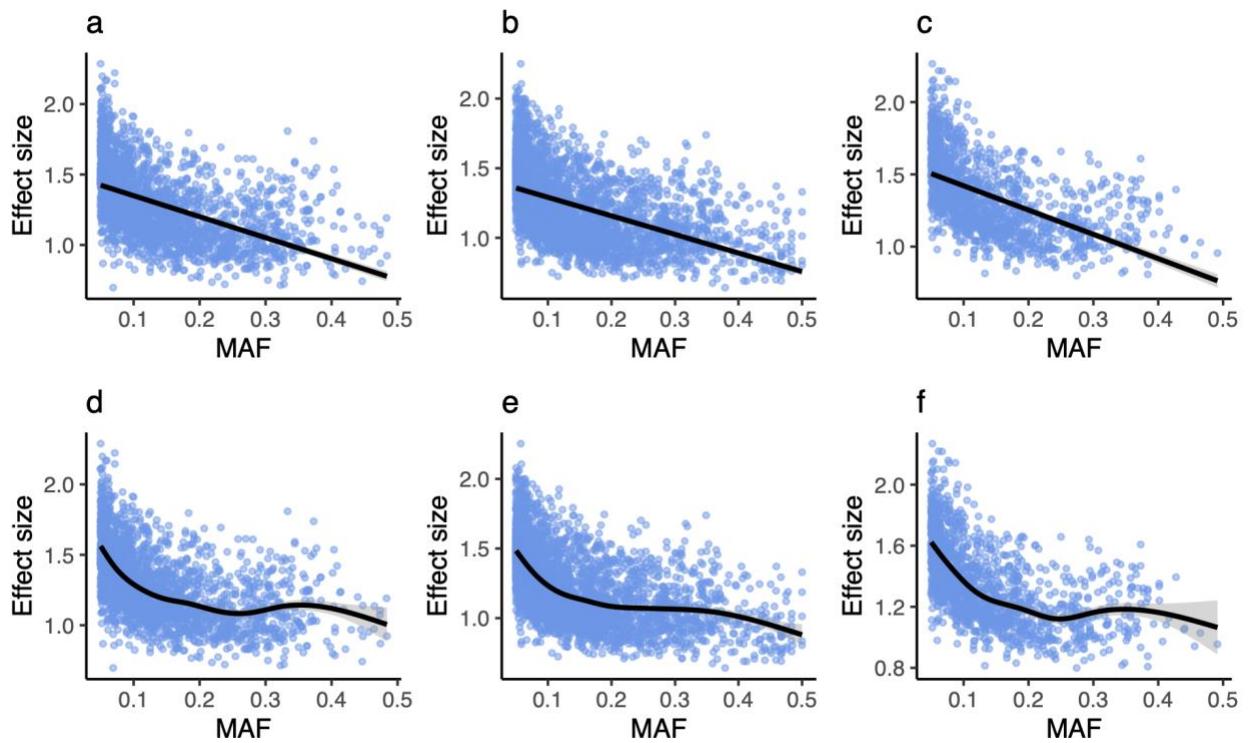
Figure A5. Correlation between effect sizes and MAFs of eQTLs in male leaf (a, d), female leaf (b, e), and pollen (c, f) of *Rumex hastatulus*. Effect size is defined as the absolute value of the slope in the linear models in eQTL mapping. a-c, linear regression, d-f: default smoothing function in R.
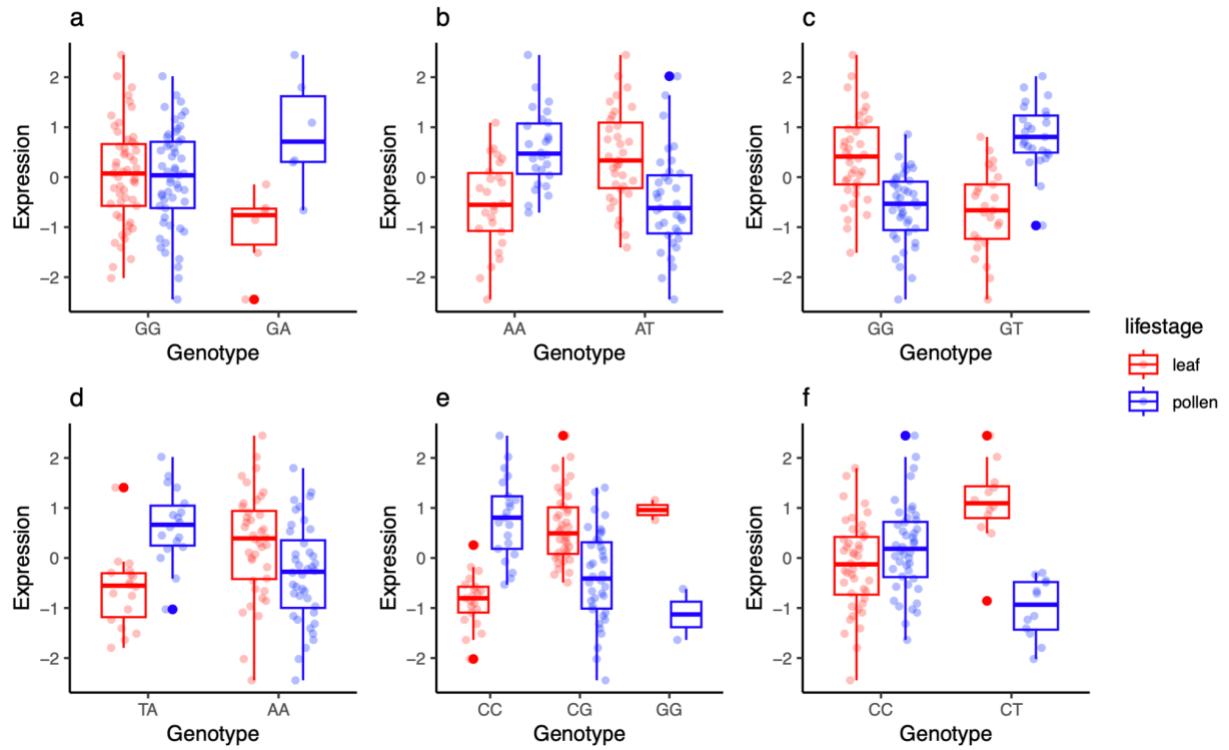
Figure A6. Examples of eQTLs showing strong discordant effects between life stages. All eQTLs are the top eQTL of the eGene in both life stages. Gene IDs and eQTL locations: a. TX_paternal_00005595 (1:255437493), b. TX_paternal_00008809 (1:455264496), c. TX_paternal_00013353 (2:162241243), d. TX_paternal_00014747 (2:238366426), e. TX_paternal_00025461 (3:13442544), f. TX_paternal_00027513 (3:75629497).
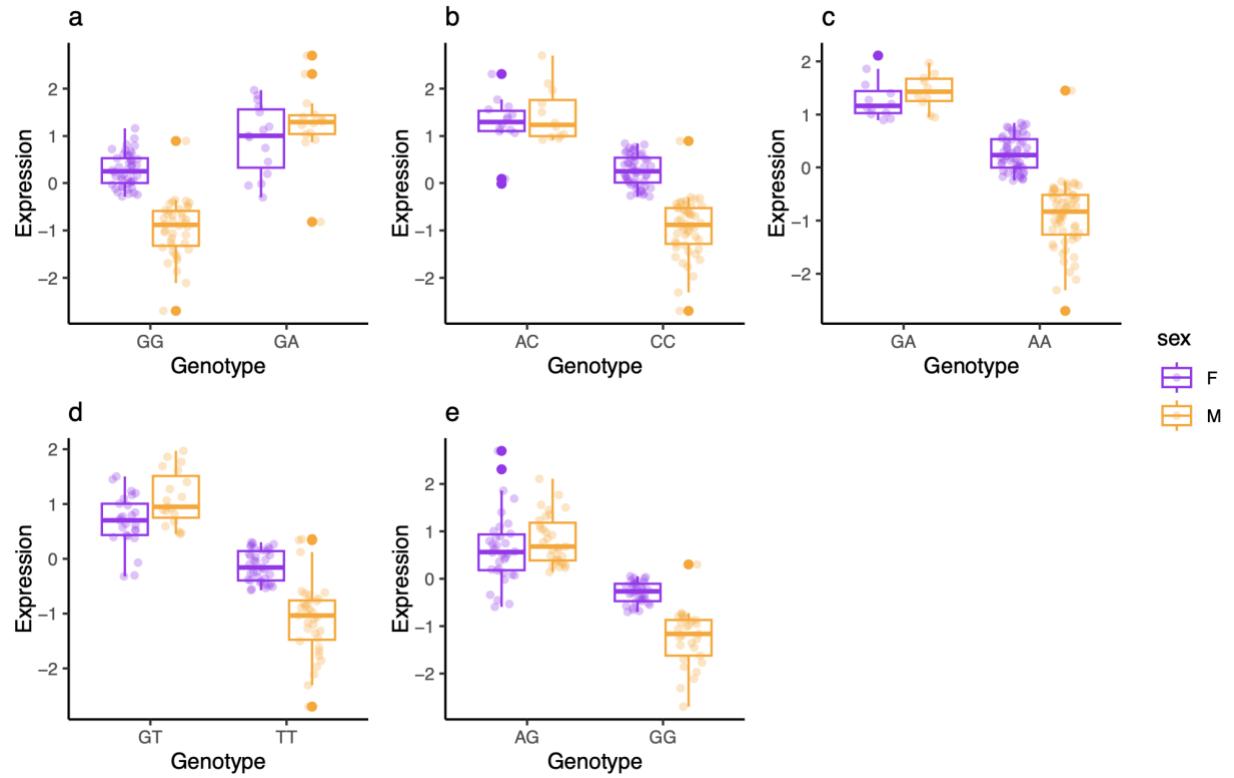
Figure A7. Examples of eQTLs showing significant Genotype × Sex interaction in *Rumex hastatulus*. Multiple-testing corrected and Benjamini-Hochberg adjusted *p*-values: 0.00197 (a), 0.00197 (b), 0.00317 (c), 0.06472 (d), 0.07921 (e). Gene IDs and eQTL locations: a. TX_paternal_00002596 (1:159008371), b. TX_paternal_00010325 (2:14033315), c. TX_paternal_00025295 (3:10654019), d. TX_paternal_00001727 (1:70371578), e. TX_paternal_00015022 (2:250749626).
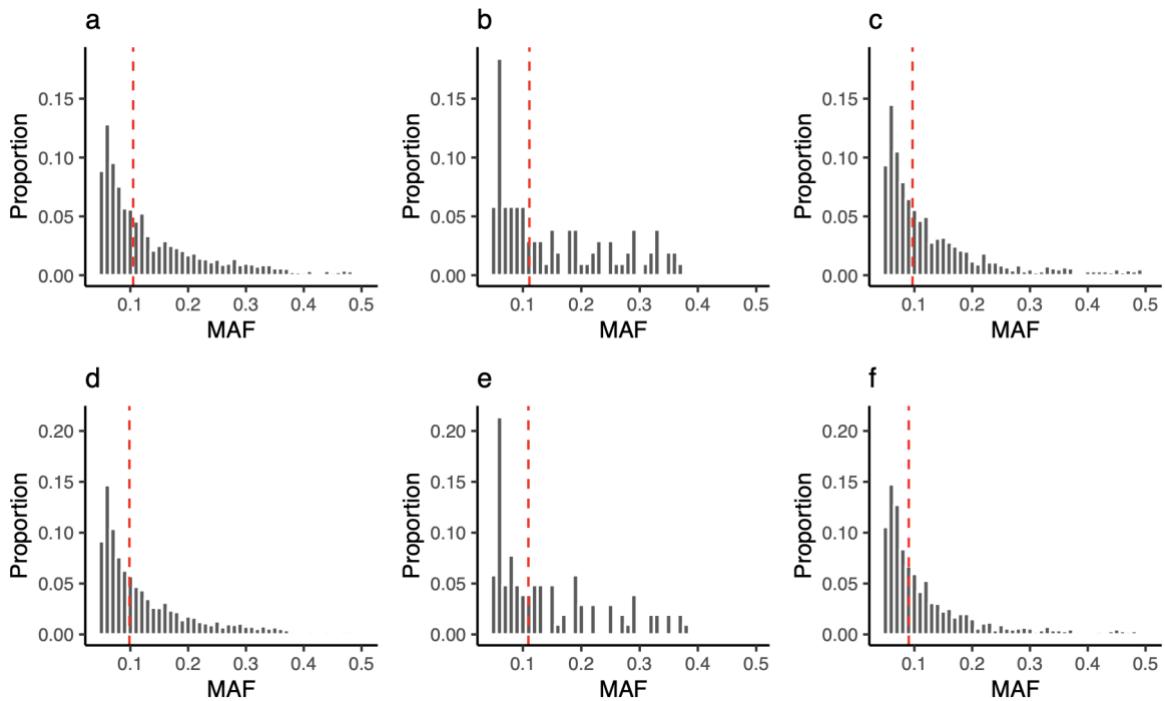
Figure A8. The distribution of MAFs of eQTLs on autosomes (a, d), PAR (b, e) and X chromosome (c, f) in female leaf of *Rumex hastatulus*. a-c: most significant eQTLs in each eGene, d-f: randomly selected eQTLs in each eGene. Binwidth = 0.01. Median of MAFs: 0.097 (a), 0.109 (b), 0.0873 (c), 0.0938 (d), 0.107 (e), 0.083 (f).
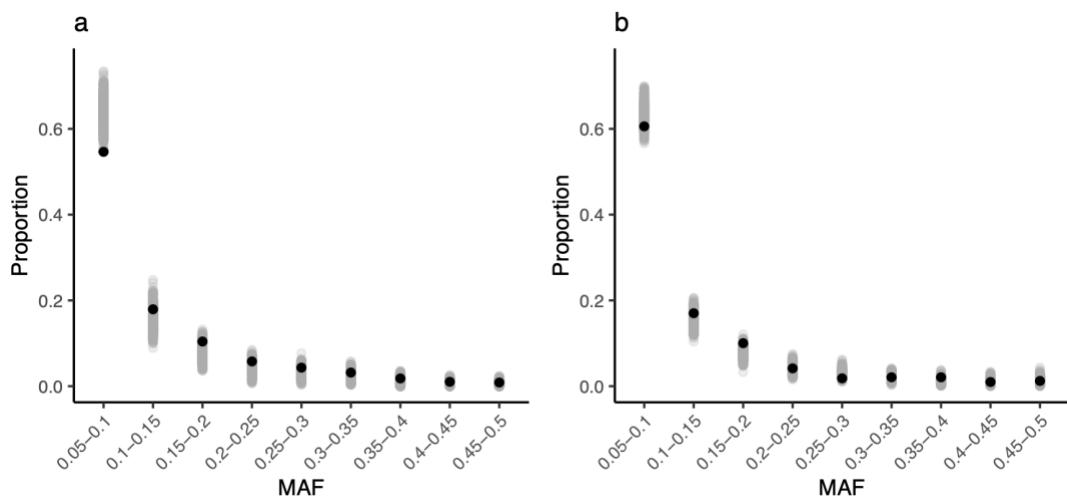


Figure A9. Distribution of MAFs of eQTLs compared to a null distribution on autosomes (a) and X-specific regions (b) in female leaf of *Rumex hastatulus*. Black dots: true positive eQTLs (observed data), grey dots: false positive eQTLs (permuted data).

## Appendix B: Chapter 3 Supplementary Materials

# Chapter 3 Supplementary Methods

*Weighting in nucleotide diversity*

We multiplied the per-site average of $\pi_s$ by the number of synonymous sites for each gene and summed the resulting $\pi_s$ and the number of synonymous sites across all genes of the gene set, we then divided the sum of $\pi_s$ and the sum of number of synonymous sites to get the weighted average $\pi_s$ for the gene set. We repeated the same steps for nonsynonymous sites to get the weighted average of $\pi_n$, and $\pi_n/\pi_s$ ratio of each gene set. We resampled the genes 1000 times with replacement within each gene set to generate the mean and 95% confidence intervals for the weighted average $\pi_s$, $\pi_n$ and $\pi_n/\pi_s$. We also calculated the mean and SEM of per-site average of $\pi_s$ and $\pi_n$ across genes for each gene set to examine the effect of weighting.

*Distribution of fitness effect (DFE) analysis*

We removed sites with missing data from the 0-fold and 4-fold VCFs used in the diversity analysis. The number of sites in different VCFs reduced to 11,046,527 (0-fold) and 2,796,108 (4-fold) in *R. hastatulus*; 16,453,490 (0-fold) and 4,503,087 (4-fold) in *C. purpureus*. At each site, *R. hastatulus* has 40 alleles from 20 female sporophytes and *C. purpureus* has 16 alleles from 8 female and 8 male gametophytes. We used the 0-fold and 4-fold VCFs for different gene sets (e.g., gametophyte- and sporophyte-specific genes) to generate the 0-fold and 4-fold site frequency spectra. We estimated the distribution of fitness effects of new mutations for different gene sets using the DFE-alpha program with default setting (Keightley and Eyre-Walker 2007). Mean and 95% confidence intervals were generated by bootstrapping the sites in different gene sets 200 times.

*Input file for balancing selection scan*

We performed whole genome alignment between *R. hastatulus* and the close relative *R. bucephalophorus* using anchorwave (Song et al. 2022). The ancestral state of each site is determined by the allele in *R. bucephalophorus*, we calculated the derived allele frequency at each site in *R. hastatulus* accordingly. We interpolated the recombination rate for each site using the recombination maps from Rifkin et al. (2022). The interpolation was conducted using either a splinefun function in the R package stats, or a linear function if the sites are out of the range of

the input recombination map
([https://github.com/tvkent/BioinformaticsUtils/blob/master/interpolate.R](https://github.com/tvkent/BioinformaticsUtils/blob/master/interpolate.R)) (Kent 2023). The derived allele frequencies in *C. purpureus* were calculated using Chile individuals (1 female, 1 male) as the outgroup (Carey et al. 2021).

## *Inversions in R. hastatulus*

Using phased genome assemblies from the XYY cytotype (Sacchi et al. 2024), and the XY cytotype of *R. hastatulus* (Humphries et al, *in prep*) we used comparative pairwise dotplots based on syntenic gene positions in COGE (Lyons and Freeling 2008) to identify large (>1MB) inversion polymorphisms between haplotypes on the autosomes. Given evidence for shared inversion polymorphisms between the cytotype populations (Sacchi et al. 2024), we recorded the locations of all inverted regions, including those that are heterozygous between haplotypes within the XYY cytotype. These regional positions were localized on the reference genome used in this assembly using COGE (Lyons and Freeling 2008) for analysis of enrichment of balancing selection signals.

# Chapter 3 Supplementary Tables

Table B1. *Rumex hastatulus* leaf DNA sequencing data.

| SampleID | Accession | Study | Biosample_accession | Sex | Read Number |
|---|---|---|---|---|---|
| 28cFLD | SRR15988666 | SRP336884 | SAMN20804014 | female | 64,585,743 |
| 42dFLD | SRR15988650 | SRP336884 | SAMN20804027 | female | 65,953,316 |
| 13bFLD | SRR15988648 | SRP336884 | SAMN20804002 | female | 66,739,339 |
| 41cFLD | SRR15988651 | SRP336884 | SAMN20804026 | female | 67,055,870 |
| 27cFLD | SRR15988677 | SRP336884 | SAMN20804013 | female | 67,319,284 |
| 1dFLD | SRR15988608 | SRP336884 | SAMN20804065 | female | 69,906,108 |
| 40aFLD | SRR15988652 | SRP336884 | SAMN20804025 | female | 71,055,865 |
| 43aFLD | SRR15988649 | SRP336884 | SAMN20804028 | female | 71,889,822 |
| 56aFLD | SRR15988635 | SRP336884 | SAMN20804040 | female | 72,156,350 |
| 55cFLD | SRR15988636 | SRP336884 | SAMN20804039 | female | 72,787,257 |
| 29cFLD | SRR15988663 | SRP336884 | SAMN20804015 | female | 73,026,787 |
| 68cFLD | SRR15988623 | SRP336884 | SAMN20804051 | female | 74,183,480 |
| 66aFLD | SRR15988624 | SRP336884 | SAMN20804050 | female | 74,662,391 |
| 69aFLD | SRR15988622 | SRP336884 | SAMN20804052 | female | 74,988,869 |
| 81cFLD | SRR15988609 | SRP336884 | SAMN20804064 | female | 75,402,838 |
| 54aFLD | SRR15988638 | SRP336884 | SAMN20804038 | female | 75,676,303 |
| 30bFLD | SRR15988662 | SRP336884 | SAMN20804016 | female | 78,606,582 |
| 2fFLD | SRR15988605 | SRP336884 | SAMN20804068 | female | 85,713,751 |
| 80bFLD | SRR15988610 | SRP336884 | SAMN20804063 | female | 86,778,222 |
| 15aFLD | SRR15988637 | SRP336884 | SAMN20804003 | female | 88,888,320 |

Table B2. *Ceratodon purpureus* sporophyte RNA sequencing data.

| Sample Name | Plate Location | LIB ID | Read Number | JGI Sequencing Project ID | FeatureCount mapped % | Used in expression analysis |
|---|---|---|---|---|---|---|
| B150_B190_G_1 | B1 | JDBL | 19827 | 1272520 | / | No |
| G150_O110_G_1 | H3 | JDCJ | 33030143 | 1272523 | 69.4% | |
| B150_O110_G_4 | D2 | JDBX | 37333146 | 1272521 | 16.1% | No |
| G150_O110_G_3 | B4 | JDCL | 43925360 | 1272524 | 64.1% | |
| B150_O110_G_2 | B2 | JDBU | 44490203 | 1272521 | 72.8% | |
| G150_B190_G_2 | D3 | JDCF | 46407871 | 1272522 | 61.4% | |
| B150_O110_G_10 | B3 | JDCD | 47877492 | 1272521 | 72.0% | |
| B150_O110_G_9 | A3 | JDCC | 48564793 | 1272521 | 72.8% | |
| G150_O110_G_2 | A4 | JDCK | 50022034 | 1272524 | 19.1% | |
| L120_O110_G_1 | C4 | JDCM | 51480420 | 1272525 | 54.9% | |
| B150_B190_G_5 | F1 | JDBQ | 51656024 | 1272520 | 66.6% | |
| G150_G100_G_1 | F3 | JDCH | 51763212 | 1272522 | 55.5% | |
| B150_B190_G_2 | C1 | JDBM | 51896988 | 1272520 | 6.9% | No |
| B150_O110_G_3 | C2 | JDBW | 52038814 | 1272521 | 43.7% | No |
| B150_O110_G_8 | H2 | JDCB | 53043265 | 1272521 | 65.1% | No |
| L120_O110_G_2 | D4 | JDCN | 53285587 | 1272525 | 37.8% | |
| B150_O110_G_6 | F2 | JDBZ | 54867152 | 1272521 | 70.4% | |
| B150_B190_G_3 | D1 | JDBN | 54931844 | 1272520 | 68.2% | |
| B150_O110_G_7 | G2 | JDCA | 55238412 | 1272521 | 42.1% | |
| G150_G100_G_2 | G3 | JDCI | 56257084 | 1272523 | 69.8% | |
| B150_O110_G_5 | E2 | JDBY | 58262420 | 1272521 | 69.7% | |
| G150_B190_G_3 | E3 | JDCG | 59420501 | 1272522 | 67.6% | |
| G150_B190_G_1 | C3 | JDCE | 59447828 | 1272521 | 63.8% | |
| B150_B190_G_4 | E1 | JDBP | 60645937 | 1272520 | 56.9% | |
| B150_O110_G_1 | A2 | JDBT | 61164338 | 1272521 | 71.6% | |
| B150_B190_G_6 | G1 | JDBR | 64278575 | 1272520 | 67.2% | |

Table B3. Number of life-stage biased or specific genes in *Rumex hastatulus* and *Ceratodon purpureus*. All genes here have a baseMean > 5 in either life stage.

***Rumex hastatulus***

| | Sporophyte-biased | | Gametophyte-biased | | Unbiased | Sporophyte-specific | | Gametophyte-specific | |
|---|---|---|---|---|---|---|---|---|---|
| | p<0.1 | p<0.1, FC >2 | p<0.1 | p<0.1, FC >2 | p>0.1 | 10% expression ratio | 10% expression ratio, FC>4 | 10% expression ratio | 10% expression ratio, FC>4 |
| **Autosome** | 8596 | 7248 | 5726 | 4744 | 714 | 1440 | 1438 | 1492 | 1490 |
| **XY** | 1705 | 1405 | 1097 | 891 | 144 | 273 | 273 | 221 | 221 |
| **Total** | 10301 | 8653 | 6823 | 5635 | 858 | 1713 | 1711 | 1713 | 1711 |

***Ceratodon purpureus***

| | Sporophyte-biased | | Gametophyte-biased | | Unbiased | Sporophyte-specific | | Gametophyte-specific | |
|---|---|---|---|---|---|---|---|---|---|
| **Autosome** | 7787 | 4394 | 8851 | 5670 | 5263 | 1687 | 1656 | 1515 | 1514 |
| **U** | 637 | 449 | 510 | 349 | 884 | 133 | 133 | 161 | 157 |
| **V** | 575 | 359 | 617 | 427 | 737 | 78 | 78 | 222 | 215 |
| **Total** | 8999 | 5202 | 9978 | 6446 | 6884 | 1898 | 1867 | 1898 | 1886 |

Table B4. GO enrichment for biased and specific genes between life stages in *Rumex hastatulus*.

| GO.ID | Term | Annotated | Significant | Expected | P_value |
|---|---|---|---|---|---|
| **Gametophyte-biased genes (log2FoldChange > 1 & padj < 0.1)** | | | | | |
| GO:0000723 | telomere maintenance | 45 | 31 | 12.02 | 3.4E-09 |
| GO:0006281 | DNA repair | 88 | 42 | 23.51 | 1.7E-08 |
| GO:0042545 | cell wall modification | 38 | 22 | 10.15 | 0.000045 |
| GO:0006887 | exocytosis | 20 | 14 | 5.34 | 0.00019 |
| GO:0006812 | cation transport | 61 | 23 | 16.3 | 0.00404 |
| GO:0043666 | regulation of phosphoprotein phosphatase activity | 4 | 4 | 1.07 | 0.00507 |
| GO:0006621 | protein retention in ER lumen | 6 | 5 | 1.6 | 0.00631 |
| GO:0007205 | protein kinase C-activating G protein-coupled receptor signaling pathway | 6 | 5 | 1.6 | 0.00631 |
| GO:0005992 | trehalose biosynthetic process | 22 | 11 | 5.88 | 0.01635 |
| GO:0006508 | proteolysis | 211 | 64 | 56.38 | 0.03026 |
| **Gametophyte-specific genes** | | | | | |
| GO.ID | Term | Annotated | Significant | Expected | P_value |
| GO:0000723 | telomere maintenance | 45 | 28 | 4.2 | 1.20E-18 |
| GO:0042545 | cell wall modification | 38 | 22 | 3.54 | 6.10E-14 |
| GO:0006281 | DNA repair | 88 | 28 | 8.2 | 1.90E-09 |
| GO:0006812 | cation transport | 61 | 13 | 5.69 | 4.30E-05 |
| GO:0055085 | transmembrane transport | 266 | 38 | 24.8 | 0.001 |
| GO:0006887 | exocytosis | 20 | 6 | 1.86 | 0.0079 |
| GO:0006468 | protein phosphorylation | 752 | 85 | 70.11 | 0.0254 |
| GO:0010215 | cellulose microfibril organization | 9 | 3 | 0.84 | 0.0441 |
| GO:0005992 | trehalose biosynthetic process | 22 | 5 | 2.05 | 0.0478 |
| GO:0000160 | phosphorelay signal transduction system | 32 | 6 | 2.98 | 0.0714 |
| **Sporophyte-biased genes (log2FoldChange > 1 & padj < 0.1)** | | | | | |
| GO.ID | Term | Annotated | Significant | Expected | P_value |
| GO:0006412 | translation | 179 | 149 | 96.78 | 8.30E-20 |
| GO:0006749 | glutathione metabolic process | 28 | 22 | 15.14 | 0.0043 |
| GO:0015979 | photosynthesis | 12 | 11 | 6.49 | 0.0072 |
| GO:0019752 | carboxylic acid metabolic process | 123 | 90 | 66.5 | 0.0085 |
| GO:0030244 | cellulose biosynthetic process | 31 | 23 | 16.76 | 0.0172 |
| GO:0016559 | peroxisome fission | 5 | 5 | 2.7 | 0.0461 |
| GO:0006075 | (1->3)-beta-D-glucan biosynthetic process | 17 | 13 | 9.19 | 0.0506 |
| GO:0032508 | DNA duplex unwinding | 8 | 7 | 4.33 | 0.0567 |
| GO:0006096 | glycolytic process | 21 | 15 | 11.35 | 0.0819 |
| GO:0009725 | response to hormone | 27 | 18 | 14.6 | 0.082 |
| **Sporophyte-specific genes** | | | | | |
| GO.ID | Term | Annotated | Significant | Expected | P_value |
| GO:0055085 | transmembrane transport | 266 | 44 | 26.54 | 1.10E-04 |
| GO:0005975 | carbohydrate metabolic process | 354 | 51 | 35.32 | 1.50E-04 |
| GO:0019752 | carboxylic acid metabolic process | 123 | 22 | 12.27 | 0.00103 |
| GO:0009116 | nucleoside metabolic process | 6 | 4 | 0.6 | 0.00124 |

| GO:0006749 | glutathione metabolic process | 28 | 8 | 2.79 | 0.0047 |
|---|---|---|---|---|---|
| GO:0006568 | tryptophan metabolic process | 2 | 2 | 0.2 | 0.00993 |
| GO:0006542 | glutamine biosynthetic process | 2 | 2 | 0.2 | 0.00993 |
| GO:0009772 | photosynthetic electron transport in photosystem II | 3 | 2 | 0.3 | 0.02782 |
| GO:0030418 | nicotianamine biosynthetic process | 3 | 2 | 0.3 | 0.02782 |
| GO:0015979 | photosynthesis | 12 | 5 | 1.2 | 0.03724 |

Table B5. GO enrichment for biased and specific genes between life stages in *Ceratodon purpureus*.

| GO.ID | Term | Annotated | Significant | Expected | P_value |
|---|---|---|---|---|---|
| **Gametophyte-biased genes (log2FoldChange > 1 & padj < 0.1)** | | | | | |
| GO:0006468 | protein phosphorylation | 938 | 323 | 232.59 | < 1e-30 |
| GO:0007165 | signal transduction | 455 | 172 | 112.82 | 2.10E-30 |
| GO:0015979 | photosynthesis | 94 | 67 | 23.31 | 5.60E-18 |
| GO:0008152 | metabolic process | 4919 | 1244 | 1219.71 | 2.00E-15 |
| GO:0009765 | photosynthesis, light harvesting | 38 | 27 | 9.42 | 2.20E-13 |
| GO:0006979 | response to oxidative stress | 83 | 38 | 20.58 | 3.70E-10 |
| GO:0009664 | plant-type cell wall organization | 28 | 20 | 6.94 | 1.60E-08 |
| GO:0048544 | recognition of pollen | 12 | 10 | 2.98 | 8.90E-07 |
| GO:0000160 | phosphorelay signal transduction system | 57 | 25 | 14.13 | 3.60E-06 |
| GO:0006508 | proteolysis | 354 | 83 | 87.78 | 4.20E-06 |
| **Gametophyte-specific genes** | | | | | |
| GO.ID | Term | Annotated | Significant | Expected | P_value |
| GO:0006468 | protein phosphorylation | 938 | 92 | 49.58 | 2.20E-18 |
| GO:0006979 | response to oxidative stress | 83 | 20 | 4.39 | 1.70E-11 |
| GO:0009664 | plant-type cell wall organization | 28 | 13 | 1.48 | 1.40E-08 |
| GO:0048544 | recognition of pollen | 12 | 7 | 0.63 | 6.70E-08 |
| GO:0008272 | sulfate transport | 6 | 4 | 0.32 | 2.80E-05 |
| GO:0042546 | cell wall biogenesis | 45 | 12 | 2.38 | 5.00E-05 |
| GO:0009765 | photosynthesis, light harvesting | 38 | 8 | 2.01 | 6.60E-05 |
| GO:0042545 | cell wall modification | 38 | 8 | 2.01 | 6.60E-05 |
| GO:0072488 | ammonium transmembrane transport | 13 | 5 | 0.69 | 7.30E-05 |
| GO:0010215 | cellulose microfibril organization | 8 | 4 | 0.42 | 0.00012 |
| **Sporophyte-biased genes (log2FoldChange > 1 & padj < 0.1)** | | | | | |
| GO.ID | Term | Annotated | Significant | Expected | P_value |
| GO:0008152 | metabolic process | 4919 | 792 | 770.92 | 1.40E-14 |
| GO:0005975 | carbohydrate metabolic process | 448 | 105 | 70.21 | 3.90E-11 |
| GO:0055085 | transmembrane transport | 415 | 93 | 65.04 | 2.20E-09 |
| GO:0007275 | multicellular organism development | 47 | 17 | 7.37 | 1.00E-06 |
| GO:0006633 | fatty acid biosynthetic process | 60 | 20 | 9.4 | 2.00E-06 |
| GO:0007017 | microtubule-based process | 108 | 19 | 16.93 | 7.00E-06 |
| GO:0006355 | regulation of DNA-templated transcription | 405 | 71 | 63.47 | 1.90E-05 |
| GO:0042545 | cell wall modification | 38 | 14 | 5.96 | 1.90E-05 |
| GO:0006465 | signal peptide processing | 6 | 5 | 0.94 | 7.60E-05 |
| GO:0009058 | biosynthetic process | 1609 | 273 | 252.17 | 8.10E-05 |
| **Sporophyte-specific genes** | | | | | |
| GO.ID | Term | Annotated | Significant | Expected | P_value |
| GO:0008152 | metabolic process | 4919 | 268 | 254.27 | 3.40E-10 |
| GO:0006633 | fatty acid biosynthetic process | 60 | 11 | 3.1 | 8.50E-06 |
| GO:0006629 | lipid metabolic process | 355 | 34 | 18.35 | 0.00012 |
| GO:0007275 | multicellular organism development | 47 | 9 | 2.43 | 0.00013 |

| GO:0055085 | transmembrane transport | 415 | 32 | 21.45 | 0.00018 |
|------------|-------------------------|-----|----|-------|---------|
| GO:0016567 | protein ubiquitination | 82 | 10 | 4.24 | 0.00072 |
| GO:0015074 | DNA integration | 2 | 2 | 0.1 | 0.00131 |
| GO:0006355 | regulation of DNA-templated transcription | 405 | 27 | 20.93 | 0.00165 |
| GO:0030001 | metal ion transport | 169 | 15 | 8.74 | 0.00232 |
| GO:0006511 | ubiquitin-dependent protein catabolic process | 67 | 8 | 3.46 | 0.00278 |

Table B6. Estimates of nucleotide diversity in bins of expression bias in *Rumex hastatulus* and *Ceratodon purpureus*.

| *Rumex hastatulus* | | | | | |
|---|---|---|---|---|---|
| log2FoldChange | Number of genes | mean_pi_syn | mean_pi_nonsyn | mean_ratio_pis | lower_pi_syn |
| (0,1.143] | 1210 | 0.008 | 0.0025 | 0.327 | 0.01 |
| (1.143,2.79] | 1130 | 0.008 | 0.0025 | 0.335 | 0.01 |
| (2.79,5.7235] | 1042 | 0.008 | 0.003 | 0.375 | 0.01 |
| (5.7235,) | 1009 | 0.009 | 0.0032 | 0.35 | 0.01 |
| (-1.255,0) | 1811 | 0.007 | 0.0023 | 0.321 | 0.01 |
| (-2.154298,-1.255] | 1847 | 0.007 | 0.0022 | 0.303 | 0.01 |
| (-3.0162,-2.154298] | 1845 | 0.007 | 0.0021 | 0.286 | 0.01 |
| (,-3.0162] | 1840 | 0.008 | 0.0024 | 0.309 | 0.01 |
| log2FoldChange | higher_pi_syn | lower_pi_nonsyn | higher_pi_nonsyn | lower_ratio_pis | higher_ratio_pis |
| (0,1.143] | 0.0081 | 0.002 | 0.0027 | 0.305 | 0.35 |
| (1.143,2.79] | 0.008 | 0.002 | 0.0027 | 0.316 | 0.36 |
| (2.79,5.7235] | 0.0083 | 0.003 | 0.0032 | 0.351 | 0.4 |
| (5.7235,) | 0.0095 | 0.003 | 0.0034 | 0.324 | 0.38 |
| (-1.255,0) | 0.0074 | 0.002 | 0.0024 | 0.306 | 0.34 |
| (-2.154298,-1.255] | 0.0075 | 0.002 | 0.0023 | 0.289 | 0.32 |
| (-3.0162,-2.154298] | 0.0078 | 0.002 | 0.0022 | 0.271 | 0.3 |
| (,-3.0162] | 0.0082 | 0.002 | 0.0026 | 0.295 | 0.32 |
| *Ceratodon purpureus* | | | | | |
| log2FoldChange | Number of genes | mean_pi_syn | mean_pi_nonsyn | mean_ratio_pis | lower_pi_syn |
| (0,0.43028] | 2564 | 0.011 | 0.003 | 0.265 | 0.01 |
| (0.43028,0.996] | 2544 | 0.013 | 0.0036 | 0.289 | 0.01 |
| (0.996,1.99] | 2517 | 0.015 | 0.0049 | 0.325 | 0.01 |
| (1.99,) | 2390 | 0.018 | 0.0064 | 0.348 | 0.02 |
| (-0.3853,0) | 2319 | 0.011 | 0.0027 | 0.251 | 0.01 |
| (-0.8416,-0.3853] | 2265 | 0.012 | 0.0027 | 0.231 | 0.01 |
| (-1.739,-0.8416] | 2073 | 0.013 | 0.0033 | 0.259 | 0.01 |
| (,-1.739] | 1943 | 0.013 | 0.0037 | 0.276 | 0.01 |
| log2FoldChange | higher_pi_syn | lower_pi_nonsyn | higher_pi_nonsyn | lower_ratio_pis | higher_ratio_pis |
| (0,0.43028] | 0.0118 | 0.003 | 0.0032 | 0.252 | 0.28 |
| (0.43028,0.996] | 0.0131 | 0.003 | 0.0039 | 0.272 | 0.3 |
| (0.996,1.99] | 0.0156 | 0.004 | 0.0052 | 0.307 | 0.34 |
| (1.99,) | 0.0191 | 0.006 | 0.0069 | 0.329 | 0.37 |
| (-0.3853,0) | 0.0113 | 0.002 | 0.0029 | 0.236 | 0.27 |
| (-0.8416,-0.3853] | 0.0123 | 0.003 | 0.0029 | 0.217 | 0.25 |
| (-1.739,-0.8416] | 0.0133 | 0.003 | 0.0035 | 0.244 | 0.28 |
| (,-1.739] | 0.014 | 0.003 | 0.0039 | 0.261 | 0.29 |

Table B7. Estimates of nucleotide diversity in different gene sets in *Rumex hastatulus* and *Ceratodon purpureus*.

| *Rumex hastatulus* | | | | | |
|---|---|---|---|---|---|
| **Tissue (N)** | **mean_pi_syn** | **mean_pi_nonsyn** | **mean_ratio_pis** | **lower_pi_syn** | **higher_pi_syn** |
| **gametophyte_specific (920)** | 0.00934467 | 0.00333246 | 0.35675475 | 0.008873542 | 0.00980046 |
| **sporophyte_specific (1149)** | 0.00814595 | 0.00254105 | 0.31205289 | 0.007728935 | 0.00856987 |
| **unbiased (615)** | 0.00796178 | 0.00279733 | 0.35167015 | 0.007369158 | 0.00861676 |
| **Tissue (N)** | lower_pi_nonsyn | higher_pi_nonsyn | lower_ratio_pis | higher_ratio_pis | |
| **gametophyte_specific (920)** | 0.0030897 | 0.00359878 | 0.32931557 | 0.387113537 | |
| **sporophyte_specific (1149)** | 0.00239277 | 0.0026969 | 0.29419947 | 0.331037308 | |
| **unbiased (615)** | 0.00258474 | 0.00303054 | 0.32216729 | 0.385513519 | |
| *Ceratodon purpureus* | | | | | |
| **Tissue (N)** | **mean_pi_syn** | **mean_pi_nonsyn** | **mean_ratio_pis** | **lower_pi_syn** | **higher_pi_syn** |
| **gametophyte_specific (1217)** | 0.0181709 | 0.00616507 | 0.33923073 | 0.017187064 | 0.01914442 |
| **sporophyte_specific (1232)** | 0.01382557 | 0.00389604 | 0.28188563 | 0.01308489 | 0.01461763 |
| **unbiased (3747)** | 0.01187295 | 0.0033327 | 0.28068688 | 0.011447877 | 0.01229793 |
| **Tissue (N)** | lower_pi_nonsyn | higher_pi_nonsyn | lower_ratio_pis | higher_ratio_pis | |
| **gametophyte_specific (1217)** | 0.00557957 | 0.00682298 | 0.31454886 | 0.364833211 | |
| **sporophyte_specific (1232)** | 0.00362908 | 0.00418478 | 0.26345267 | 0.299689003 | |
| **unbiased (3747)** | 0.00314225 | 0.00353279 | 0.26798292 | 0.293819751 | |

Table B8. DFE estimates of lifestage-specific genes in *Rumex hastatulus* and *Ceratodon purpureus*.

| *Rumex hastatulus* | | | | | | |
|---|---|---|---|---|---|---|
| **Category** | **Nes** | **mean** | **sd** | **sem** | **2.5%** | **97.5%** |
| gametophyte | 0~1 | 0.35325108 | 0.014295 | 0.00101081 | 0.32030955 | 0.37439683 |
| gametophyte | 1~10 | 0.1179768 | 0.01848577 | 0.00130714 | 0.0935136 | 0.16360463 |
| gametophyte | >10 | 0.5287721 | 0.00919887 | 0.00065046 | 0.50931048 | 0.5451206 |
| sporophyte | 0~1 | 0.27563035 | 0.00798238 | 0.00056444 | 0.26107178 | 0.29145268 |
| sporophyte | 1~10 | 0.13944916 | 0.00924776 | 0.00065392 | 0.12211788 | 0.15575478 |
| sporophyte | >10 | 0.58492045 | 0.00714176 | 0.000505 | 0.5714254 | 0.59800215 |
| unbiased | 0~1 | 0.32230561 | 0.01752797 | 0.00123941 | 0.2882271 | 0.34983018 |
| unbiased | 1~10 | 0.15772805 | 0.02551747 | 0.00180436 | 0.11965623 | 0.20686623 |
| unbiased | >10 | 0.51996636 | 0.01290509 | 0.00091253 | 0.4938404 | 0.54302268 |
| *Ceratodon purpureus* | | | | | | |
| **Category** | **Nes** | **mean** | **sd** | **sem** | **2.5%** | **97.5%** |
| gametophyte | 0~1 | 0.25060844 | 0.00496302 | 0.00035094 | 0.24007008 | 0.25920293 |
| gametophyte | 1~10 | 0.16064914 | 0.00883106 | 0.00062445 | 0.14555033 | 0.18123753 |
| gametophyte | >10 | 0.58874247 | 0.00679518 | 0.00048049 | 0.57305105 | 0.60226055 |
| sporophyte | 0~1 | 0.21823385 | 0.00763121 | 0.00053961 | 0.20444788 | 0.23283618 |
| sporophyte | 1~10 | 0.19724023 | 0.01493159 | 0.00105582 | 0.17170965 | 0.22289788 |
| sporophyte | >10 | 0.58452585 | 0.01008605 | 0.00071319 | 0.56633438 | 0.60232378 |
| unbiased | 0~1 | 0.19419456 | 0.00241485 | 0.00017076 | 0.18923478 | 0.19871978 |
| unbiased | 1~10 | 0.18458626 | 0.00450884 | 0.00031882 | 0.17616045 | 0.19235183 |
| unbiased | >10 | 0.62121918 | 0.00376081 | 0.00026593 | 0.61441528 | 0.62854845 |

Table B9. DFE estimates of lifestage-specific genes across expression level quantiles in *Rumex hastatulus* and *Ceratodon purpureus*.

| Rumex hastatulus | | | | | | | |
|---|---|---|---|---|---|---|---|
| Category | Nes | mean | sd | sem | 2.5% | 97.5% | Quantile |
| gametophyte | 0~1 | 0.6013073 | 0.03553812 | 0.00251292 | 0.5315169 | 0.65911393 | qt1 |
| gametophyte | 1~10 | 0.12767128 | 0.05232698 | 0.00370008 | 0.07460065 | 0.26199043 | qt1 |
| gametophyte | >10 | 0.2710214 | 0.03126545 | 0.0022108 | 0.19605733 | 0.31575628 | qt1 |
| sporophyte | 0~1 | 0.42250825 | 0.03664302 | 0.00259105 | 0.35292063 | 0.4935855 | qt1 |
| sporophyte | 1~10 | 0.15991176 | 0.04809448 | 0.00340079 | 0.0738167 | 0.26271335 | qt1 |
| sporophyte | >10 | 0.41758005 | 0.03307028 | 0.00233842 | 0.33882405 | 0.47423253 | qt1 |
| unbiased | 0~1 | 0.37755429 | 0.03024792 | 0.00213885 | 0.32153603 | 0.4328712 | qt1 |
| unbiased | 1~10 | 0.21124695 | 0.04414496 | 0.00312152 | 0.13268353 | 0.30189523 | qt1 |
| unbiased | >10 | 0.41119876 | 0.02652666 | 0.00187572 | 0.3625241 | 0.45483088 | qt1 |
| gametophyte | 0~1 | 0.47250633 | 0.03411287 | 0.00241214 | 0.39096378 | 0.52605163 | qt2 |
| gametophyte | 1~10 | 0.11405868 | 0.04349929 | 0.00307586 | 0.06028278 | 0.2194268 | qt2 |
| gametophyte | >10 | 0.41343499 | 0.02768175 | 0.0019574 | 0.3598209 | 0.4658022 | qt2 |
| sporophyte | 0~1 | 0.30798918 | 0.02014211 | 0.00142426 | 0.27027128 | 0.34709433 | qt2 |
| sporophyte | 1~10 | 0.14514453 | 0.02679324 | 0.00189457 | 0.09503548 | 0.20127188 | qt2 |
| sporophyte | >10 | 0.5468663 | 0.01674486 | 0.00118404 | 0.51237575 | 0.57678563 | qt2 |
| unbiased | 0~1 | 0.39018272 | 0.03011831 | 0.00212969 | 0.33105458 | 0.4453477 | qt2 |
| unbiased | 1~10 | 0.14389425 | 0.03896212 | 0.00275504 | 0.07069955 | 0.21770505 | qt2 |
| unbiased | >10 | 0.46592297 | 0.02455035 | 0.00173597 | 0.41850843 | 0.51443505 | qt2 |
| gametophyte | 0~1 | 0.34479259 | 0.03373562 | 0.00238547 | 0.27914435 | 0.40916068 | qt3 |
| gametophyte | 1~10 | 0.19719792 | 0.05240224 | 0.0037054 | 0.10868513 | 0.3172265 | qt3 |
| gametophyte | >10 | 0.45800949 | 0.03323518 | 0.00235008 | 0.3855936 | 0.51009885 | qt3 |
| sporophyte | 0~1 | 0.27705289 | 0.01902197 | 0.00134506 | 0.2376827 | 0.31069718 | qt3 |
| sporophyte | 1~10 | 0.16616114 | 0.0219449 | 0.00155174 | 0.12865243 | 0.21156065 | qt3 |
| sporophyte | >10 | 0.55678598 | 0.01554364 | 0.0010991 | 0.52808598 | 0.58518718 | qt3 |
| unbiased | 0~1 | 0.26908192 | 0.02330155 | 0.00164767 | 0.23092938 | 0.3165328 | qt3 |
| unbiased | 1~10 | 0.13675598 | 0.03090924 | 0.00218561 | 0.07008308 | 0.18803798 | qt3 |
| unbiased | >10 | 0.59416214 | 0.01929425 | 0.00136431 | 0.55734538 | 0.62822108 | qt3 |
| gametophyte | 0~1 | 0.19166509 | 0.01006537 | 0.00071173 | 0.17542698 | 0.21541165 | qt4 |
| gametophyte | 1~10 | 0.12094061 | 0.01027516 | 0.00072656 | 0.10184458 | 0.14237028 | qt4 |
| gametophyte | >10 | 0.68739432 | 0.00756036 | 0.0005346 | 0.67185613 | 0.70212013 | qt4 |
| sporophyte | 0~1 | 0.21631816 | 0.01078693 | 0.00076275 | 0.19395558 | 0.2373967 | qt4 |
| sporophyte | 1~10 | 0.1131724 | 0.01212033 | 0.00085704 | 0.09309223 | 0.1415968 | qt4 |
| sporophyte | >10 | 0.67050944 | 0.0084436 | 0.00059705 | 0.65378118 | 0.68660383 | qt4 |
| unbiased | 0~1 | 0.19932299 | 0.02079114 | 0.00147016 | 0.1628325 | 0.23675005 | qt4 |
| unbiased | 1~10 | 0.12456261 | 0.02433936 | 0.00172105 | 0.07831035 | 0.17416738 | qt4 |
| unbiased | >10 | 0.67611453 | 0.0157864 | 0.00111627 | 0.64731245 | 0.7062306 | qt4 |
| Ceratodon purpureus | | | | | | | |
| Category | Nes | mean | sd | sem | 2.5% | 97.5% | Quantile |
| gametophyte | 0~1 | 0.35555825 | 0.02197033 | 0.00155354 | 0.31185295 | 0.39283085 | qt1 |
| gametophyte | 1~10 | 0.3149427 | 0.05160893 | 0.0036493 | 0.23226818 | 0.430459 | qt1 |
| gametophyte | >10 | 0.32949909 | 0.03697926 | 0.00261483 | 0.25157715 | 0.38781803 | qt1 |

| sporophyte | 0~1 | 0.30403728 | 0.01475945 | 0.00104365 | 0.27577623 | 0.32986725 | qt1 |
|------------|-----|------------|------------|------------|------------|------------|-----|
| sporophyte | 1~10 | 0.24379904 | 0.03312227 | 0.0023421 | 0.18413715 | 0.3108275 | qt1 |
| sporophyte | >10 | 0.45216371 | 0.02349029 | 0.00166101 | 0.40063535 | 0.49503133 | qt1 |
| unbiased | 0~1 | 0.44259411 | 0.02039045 | 0.00144182 | 0.4076094 | 0.48209033 | qt1 |
| unbiased | 1~10 | 0.27816344 | 0.05053068 | 0.00357306 | 0.1986432 | 0.37916968 | qt1 |
| unbiased | >10 | 0.27924245 | 0.03705243 | 0.00262 | 0.20003195 | 0.34140258 | qt1 |
| gametophyte | 0~1 | 0.28550026 | 0.01140153 | 0.00080621 | 0.2651166 | 0.30860238 | qt2 |
| gametophyte | 1~10 | 0.22745225 | 0.02558538 | 0.00180916 | 0.18447635 | 0.2819531 | qt2 |
| gametophyte | >10 | 0.48704749 | 0.01869655 | 0.00132205 | 0.4467646 | 0.5163898 | qt2 |
| sporophyte | 0~1 | 0.19056417 | 0.0103663 | 0.00073301 | 0.1701879 | 0.21294763 | qt2 |
| sporophyte | 1~10 | 0.25167822 | 0.02407789 | 0.00170256 | 0.21094483 | 0.30167388 | qt2 |
| sporophyte | >10 | 0.55775764 | 0.01810483 | 0.0012802 | 0.5206043 | 0.5887295 | qt2 |
| unbiased | 0~1 | 0.27140672 | 0.01037 | 0.00073327 | 0.25232695 | 0.2921302 | qt2 |
| unbiased | 1~10 | 0.30589151 | 0.02655758 | 0.0018779 | 0.26091178 | 0.35237753 | qt2 |
| unbiased | >10 | 0.42270177 | 0.01916092 | 0.00135488 | 0.38187125 | 0.45463178 | qt2 |
| gametophyte | 0~1 | 0.22145025 | 0.00989129 | 0.00069942 | 0.20440668 | 0.23991378 | qt3 |
| gametophyte | 1~10 | 0.1475491 | 0.01729035 | 0.00122261 | 0.11745745 | 0.17960695 | qt3 |
| gametophyte | >10 | 0.63100074 | 0.0123343 | 0.00087217 | 0.61079935 | 0.65670683 | qt3 |
| sporophyte | 0~1 | 0.17789806 | 0.01163228 | 0.00082253 | 0.15525488 | 0.19845795 | qt3 |
| sporophyte | 1~10 | 0.15433544 | 0.02126868 | 0.00150392 | 0.1198696 | 0.20660788 | qt3 |
| sporophyte | >10 | 0.6677665 | 0.01484649 | 0.00104981 | 0.63504818 | 0.69317613 | qt3 |
| unbiased | 0~1 | 0.17402873 | 0.00394299 | 0.00027881 | 0.16602223 | 0.18177475 | qt3 |
| unbiased | 1~10 | 0.14951416 | 0.00663546 | 0.0004692 | 0.1368424 | 0.16212613 | qt3 |
| unbiased | >10 | 0.67645716 | 0.00545823 | 0.00038596 | 0.6660034 | 0.68710743 | qt3 |
| gametophyte | 0~1 | 0.14864621 | 0.00674647 | 0.00047705 | 0.13428748 | 0.15988405 | qt4 |
| gametophyte | 1~10 | 0.06548356 | 0.01152575 | 0.00081499 | 0.04362065 | 0.0862094 | qt4 |
| gametophyte | >10 | 0.78587027 | 0.00833272 | 0.00058921 | 0.7708551 | 0.803638 | qt4 |
| sporophyte | 0~1 | 0.12794673 | 0.00979395 | 0.00069254 | 0.11139058 | 0.14758875 | qt4 |
| sporophyte | 1~10 | 0.10737016 | 0.01663719 | 0.00117643 | 0.0720925 | 0.1378535 | qt4 |
| sporophyte | >10 | 0.7646831 | 0.01234597 | 0.00087299 | 0.74215928 | 0.7877831 | qt4 |
| unbiased | 0~1 | 0.09741431 | 0.01213349 | 0.00085797 | 0.08372 | 0.11764843 | qt4 |
| unbiased | 1~10 | 0.13244587 | 0.02095029 | 0.00148141 | 0.09896068 | 0.15558263 | qt4 |
| unbiased | >10 | 0.7701398 | 0.00949797 | 0.00067161 | 0.7571011 | 0.78740873 | qt4 |

Table B10. Number of biased genes under balancing selection vs across the whole genome. Criteria for biased expression: padj < 0.1.

| Rumex hastatulus | | | |
|---|---|---|---|
| category | Gametophyte-biased | Unbiased | Sporophyte-biased |
| balancing_sel | 142 | 21 | 242 |
| whole_genome | 6823 | 822 | 10301 |
| **Ceratodon purpureus** | | | |
| **category** | **Gametophyte-biased** | **Unbiased** | **Sporophyte-biased** |
| balancing_sel | 334 | 193 | 280 |
| whole_genome | 8851 | 5263 | 7787 |

Table B11. GO enrichment for genes under balancing selection in *Rumex hastatulus*.

| **All expressed genes under balancing selection** | | | | | |
|---|---|---|---|---|---|
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P_value** |
| GO:0032508 | DNA duplex unwinding | 8 | 2 | 0.21 | 0.017 |
| GO:0007021 | tubulin complex assembly | 1 | 1 | 0.03 | 0.026 |
| GO:0007023 | post-chaperonin tubulin folding pathway | 1 | 1 | 0.03 | 0.026 |
| GO:0006468 | protein phosphorylation | 752 | 28 | 19.88 | 0.029 |
| GO:0034220 | ion transmembrane transport | 29 | 3 | 0.77 | 0.076 |
| GO:0030418 | nicotianamine biosynthetic process | 3 | 1 | 0.08 | 0.077 |
| GO:0006891 | intra-Golgi vesicle-mediated transport | 3 | 1 | 0.08 | 0.077 |
| GO:0006486 | protein glycosylation | 39 | 3 | 1.03 | 0.083 |
| GO:0006897 | endocytosis | 4 | 1 | 0.11 | 0.102 |
| GO:0016192 | vesicle-mediated transport | 106 | 7 | 2.8 | 0.118 |
| **All expressed genes under balancing selection after excluding inversions** | | | | | |
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P_value** |
| GO:0032508 | DNA duplex unwinding | 8 | 2 | 0.15 | 0.0085 |
| GO:0007021 | tubulin complex assembly | 1 | 1 | 0.02 | 0.0182 |
| GO:0007023 | post-chaperonin tubulin folding pathway | 1 | 1 | 0.02 | 0.0182 |
| GO:0006468 | protein phosphorylation | 752 | 21 | 13.68 | 0.0215 |
| GO:0006486 | protein glycosylation | 39 | 3 | 0.71 | 0.033 |
| GO:0030418 | nicotianamine biosynthetic process | 3 | 1 | 0.05 | 0.0536 |
| GO:0006891 | intra-Golgi vesicle-mediated transport | 3 | 1 | 0.05 | 0.0536 |
| GO:0006897 | endocytosis | 4 | 1 | 0.07 | 0.0708 |
| GO:0006368 | transcription elongation by RNA polymerase II | 6 | 1 | 0.11 | 0.1044 |
| GO:0016192 | vesicle-mediated transport | 106 | 5 | 1.93 | 0.1394 |
| **Gametophyte-biased genes under balancing selection** | | | | | |
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P_value** |
| GO:0006468 | protein phosphorylation | 241 | 12 | 5.24 | 0.0021 |
| GO:0006368 | transcription elongation by RNA polymerase II | 3 | 1 | 0.07 | 0.0639 |
| GO:0006821 | chloride transport | 4 | 1 | 0.09 | 0.0843 |
| GO:0016042 | lipid catabolic process | 5 | 1 | 0.11 | 0.1043 |
| GO:0007205 | protein kinase C-activating G protein-coupled receptor signaling pathway | 6 | 1 | 0.13 | 0.1239 |
| GO:0008033 | tRNA processing | 8 | 1 | 0.17 | 0.1618 |
| GO:1902600 | proton transmembrane transport | 12 | 1 | 0.26 | 0.233 |
| GO:0006364 | rRNA processing | 13 | 1 | 0.28 | 0.2499 |
| GO:0035556 | intracellular signal transduction | 16 | 1 | 0.35 | 0.2983 |
| GO:0048193 | Golgi vesicle transport | 17 | 1 | 0.37 | 0.3138 |
| **Gametophyte-biased genes under balancing selection after excluding inversions** | | | | | |
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P_value** |
| GO:0006468 | protein phosphorylation | 241 | 8 | 3.15 | 0.0055 |
| GO:0006368 | transcription elongation by RNA polymerase II | 3 | 1 | 0.04 | 0.0387 |
| GO:0008033 | tRNA processing | 8 | 1 | 0.1 | 0.1001 |
| GO:0006355 | regulation of DNA-templated transcription | 102 | 3 | 1.33 | 0.1414 |

| GO:0006364 | rRNA processing | 13 | 1 | 0.17 | 0.1578 |
| GO:0007018 | microtubule-based movement | 19 | 1 | 0.25 | 0.2225 |
| GO:0022613 | ribonucleoprotein complex biogenesis | 16 | 1 | 0.21 | 1 |
| GO:0034660 | ncRNA metabolic process | 27 | 2 | 0.35 | 1 |
| GO:0009889 | regulation of biosynthetic process | 102 | 3 | 1.33 | 1 |
| GO:0007017 | microtubule-based process | 22 | 1 | 0.29 | 1 |

Table B12. GO enrichment for genes under balancing selection in *Ceratodon purpureus*.

| GO.ID | Term | Annotated | Significant | Expected | P_value |
|---|---|---|---|---|---|
| **All expressed genes under balancing selection** | | | | | |
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P_value** |
| GO:0006468 | protein phosphorylation | 938 | 41 | 36.73 | 0.0011 |
| GO:0055085 | transmembrane transport | 415 | 25 | 16.25 | 0.0024 |
| GO:0005986 | sucrose biosynthetic process | 4 | 2 | 0.16 | 0.0037 |
| GO:0072488 | ammonium transmembrane transport | 13 | 3 | 0.51 | 0.0039 |
| GO:0006812 | cation transport | 209 | 15 | 8.18 | 0.0042 |
| GO:0019953 | sexual reproduction | 7 | 2 | 0.27 | 0.0125 |
| GO:0006855 | xenobiotic transmembrane transport | 23 | 3 | 0.9 | 0.0198 |
| GO:0007264 | small GTPase-mediated signal transduction | 74 | 6 | 2.9 | 0.0243 |
| GO:0000271 | polysaccharide biosynthetic process | 26 | 3 | 1.02 | 0.0253 |
| GO:0035434 | copper ion transmembrane transport | 1 | 1 | 0.04 | 0.0255 |
| **Gametophyte-biased genes under balancing selection** | | | | | |
| **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **P_value** |
| GO:0042128 | nitrate assimilation | 3 | 1 | 0.06 | 0.044 |
| GO:0007094 | mitotic spindle assembly checkpoint signaling | 3 | 1 | 0.06 | 0.044 |
| GO:0016226 | iron-sulfur cluster assembly | 4 | 1 | 0.07 | 0.058 |
| GO:0055085 | transmembrane transport | 18 | 1 | 0.33 | 0.242 |
| GO:0031163 | metallo-sulfur cluster assembly | 4 | 1 | 0.07 | 1 |
| GO:2001251 | negative regulation of chromosome organization | 3 | 1 | 0.06 | 1 |
| GO:0051726 | regulation of cell cycle | 3 | 1 | 0.06 | 1 |
| GO:0008150 | biological_process | 187 | 4 | 3.46 | 1 |
| GO:0008152 | metabolic process | 105 | 2 | 1.94 | 1 |
| GO:0010639 | negative regulation of organelle organization | 3 | 1 | 0.06 | 1 |

# Chapter 3 Supplementary Figures
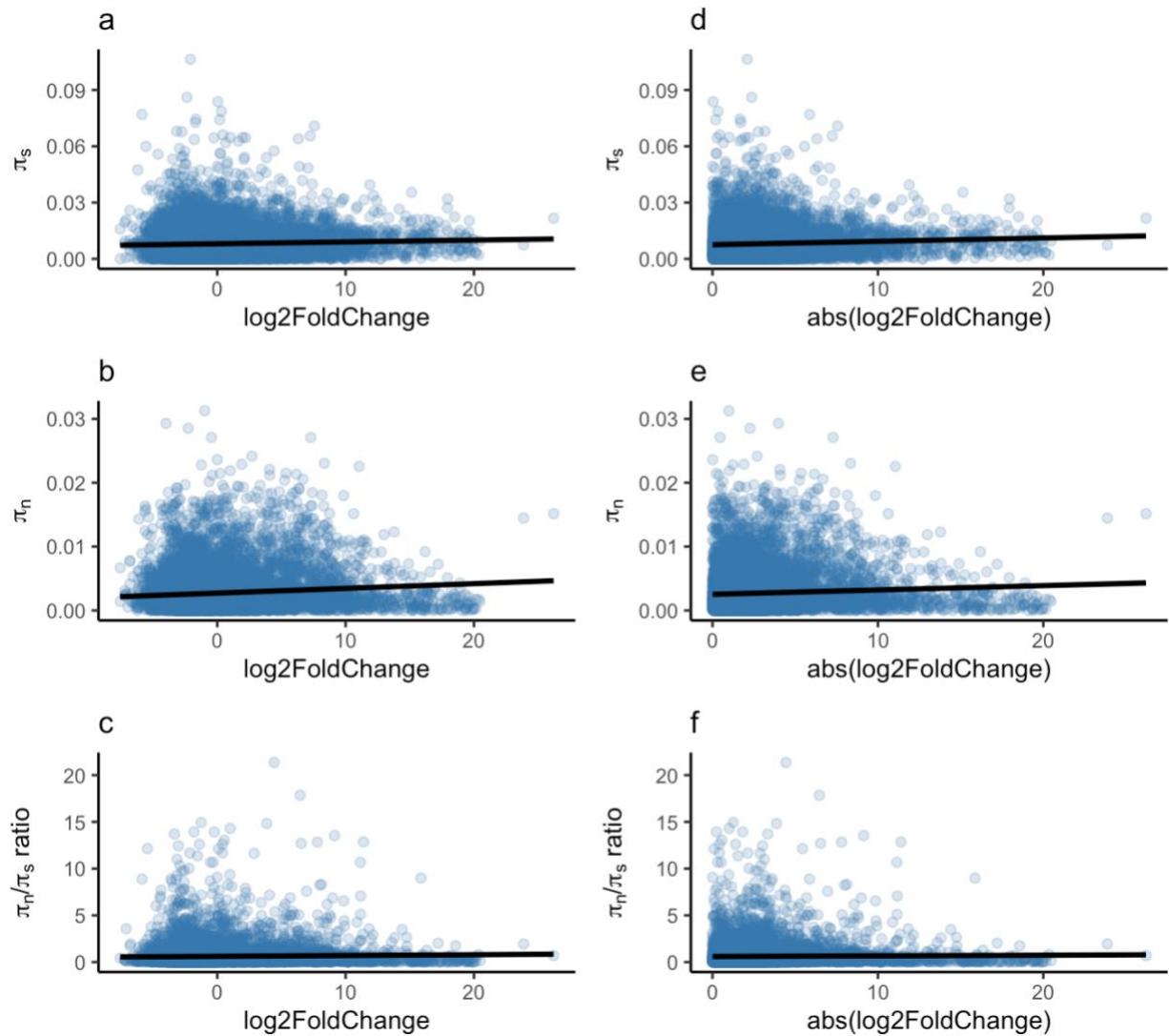


Figure B1. Linear regression between expression bias and nucleotide diversity in *R. hastatulus*. x-axis: log2FoldChange (a-c) or the absolute value of log2FoldChange (d-f). We removed zeros for $\pi_s$ and $\pi_n$ to calculate the $\pi_n/\pi_s$ ratio for each gene. The slopes and p-values for $\pi$ are all < 0.001 and < 10-8, respectively. The slopes for the $\pi_n/\pi_s$ ratio are both < 0.01, p-value = 0.00133 (c), 0.0766 (f).
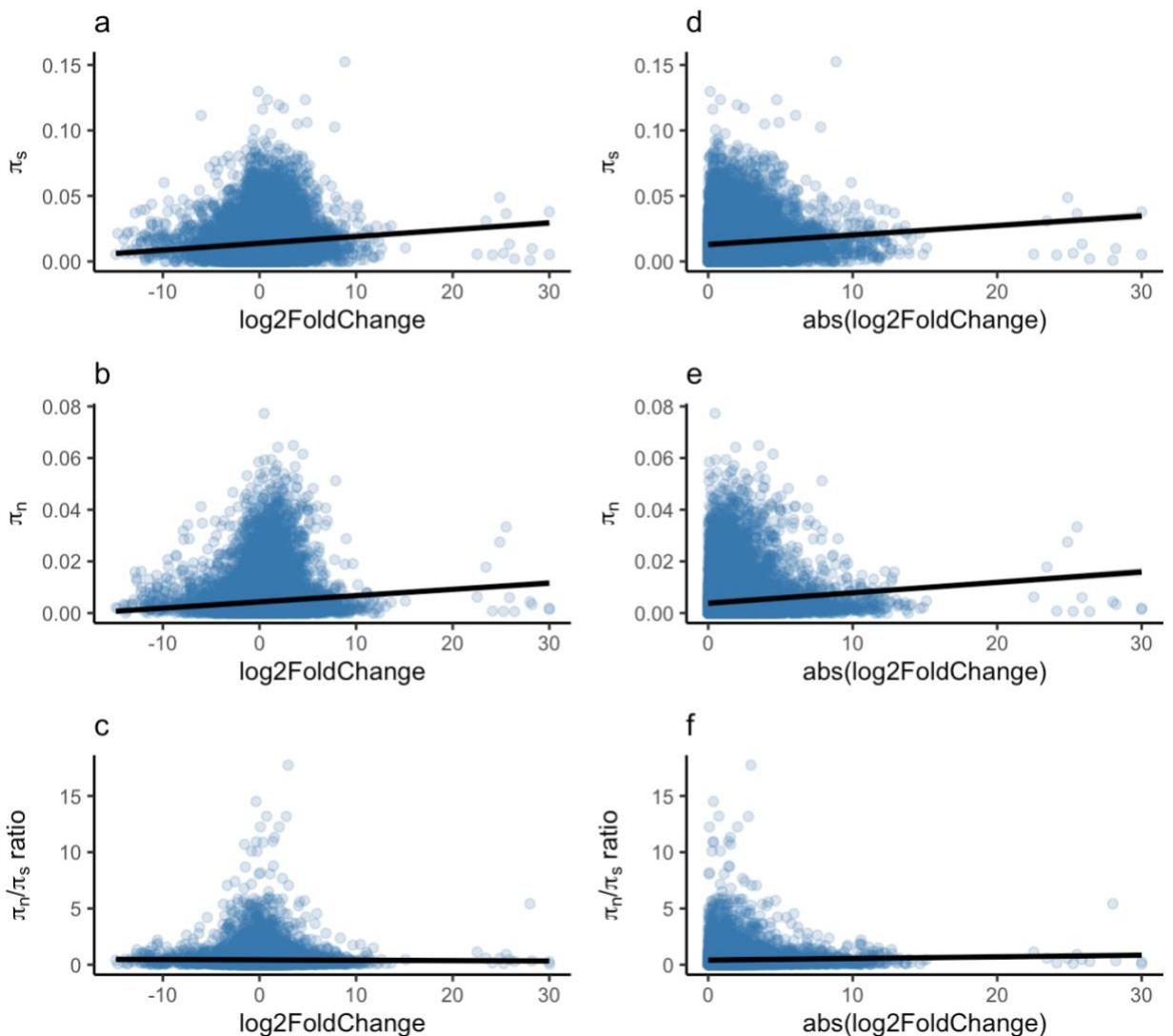
Figure B2. Linear regression between expression bias and nucleotide diversity in *C. purpureus*. x-axis: log2FoldChange (a-c) or the absolute value of log2FoldChange (d-f). We removed zeros for $\pi_s$ and $\pi_n$ to calculate the $\pi_n/\pi_s$ ratio for each gene. The slopes and p-values for $\pi$ are all < 0.001 and < 10-16, respectively. For the $\pi_n/\pi_s$ ratio: slope = -0.003242, p-value = 0.176 (c); slope = 0.014604, p-value = 1.73e-06 (f).
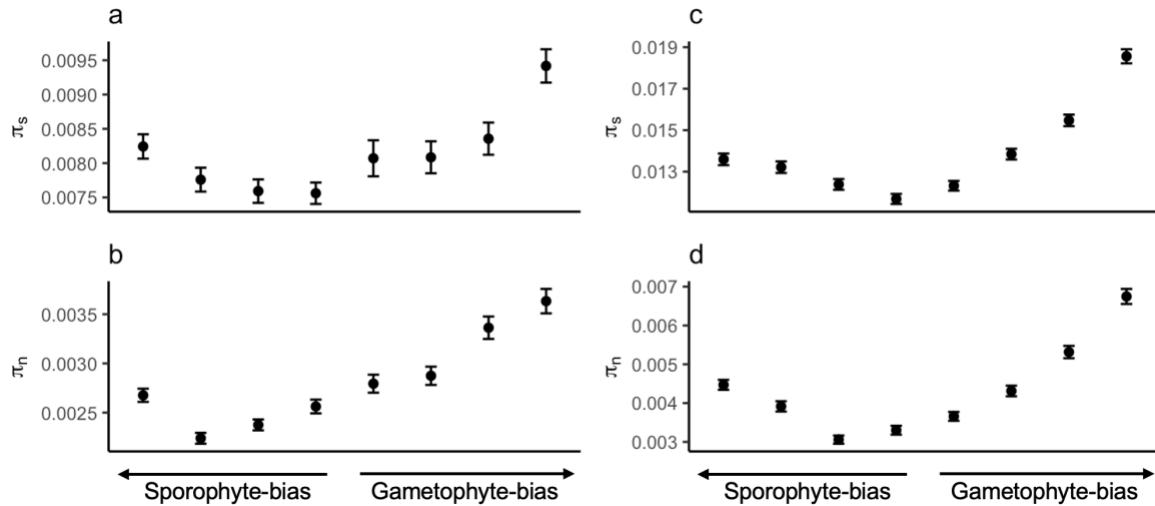
Figure B3. Effect of expression bias between life stages on nucleotide diversity in *R. hastatulus* (a-b) and *C. purpureus* (c-d). The numbers of genes in each bin are the same as Figure 1. Error bars represent mean and SEM across genes in each bin. Note that the scales on the y-axis are different between species.
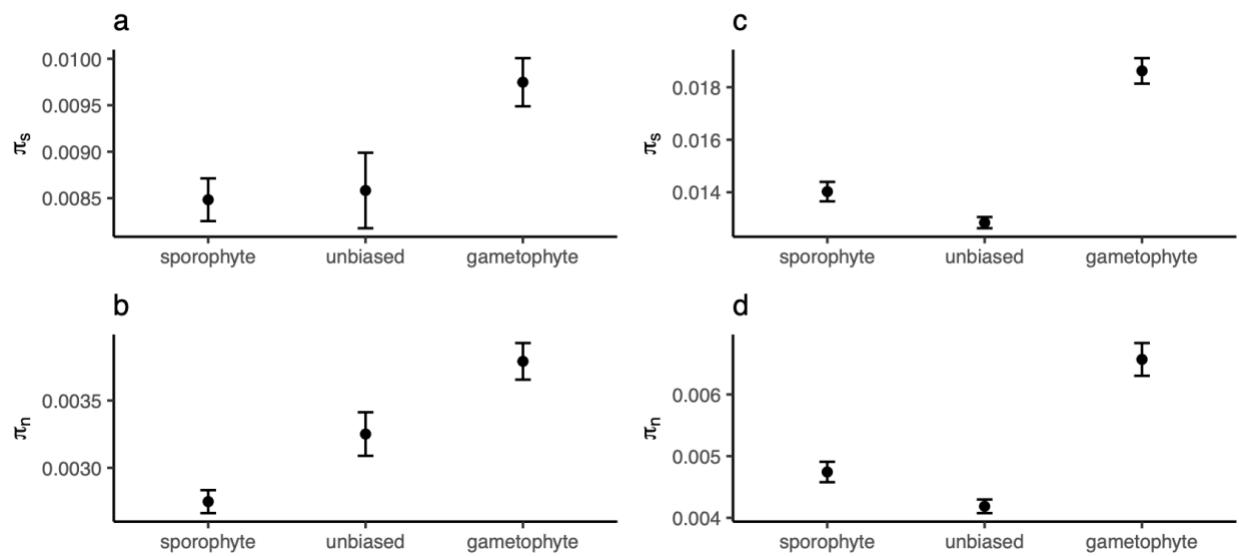


Figure B4. Mean nucleotide diversity of gametophyte-specific, sporophyte-specific, unbiased genes in *R. hastatulus* (a-b) and *C. purpureus* (c-d). The numbers of genes in each category are the same as Figure 2. Error bars represent mean and SEM across genes in each category. Note that the scales on the y-axis are different between species.
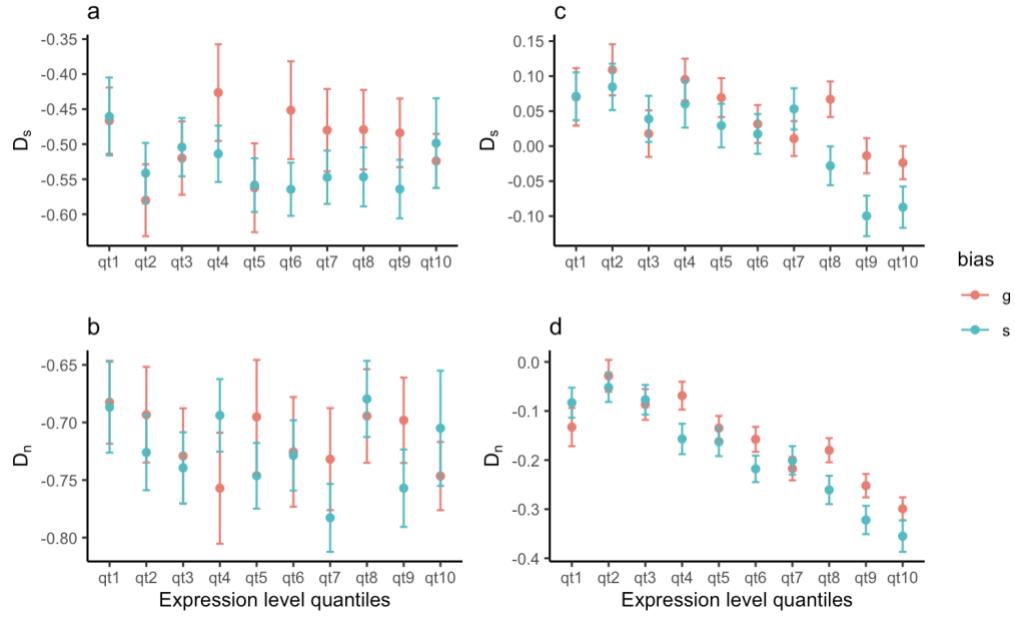
Figure B5. Effect of expression bias between life stages and expression level on Tajima's $D$. Number of genes in each bin: 676-843 ($D_s$), 1,150-1,339 ($D_n$) for *R. hastatulus* (a, b); 1,265-2,250 ($D_s$), 1,456-2,090 ($D_n$) for *C. purpureus* (c, d). qt1: lowest expression level, qt10: highest expression level. Error bars represent mean and SEM in each bin. Note that the scales on the y-axis are different between species.
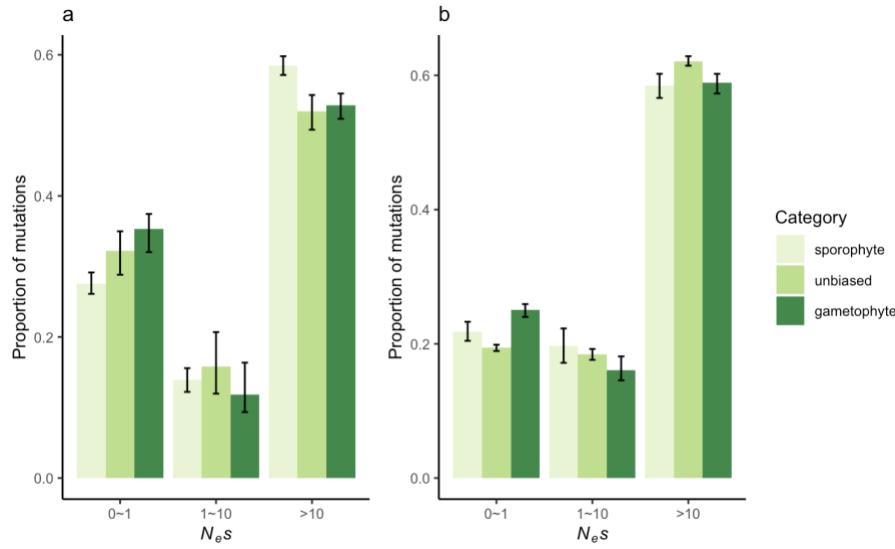
Figure B6. Distribution of fitness effects of life stage-specific genes in *R. hastatulus* (a) and *C. purpureus* (b). Number of genes in each category: 1771 (g), 1771(s) for *R. hastatulus* 1514 (g), 1656 (s) for *C. purpureus*. The mean and 95% confidence intervals are based on 200 bootstraps of the original dataset.
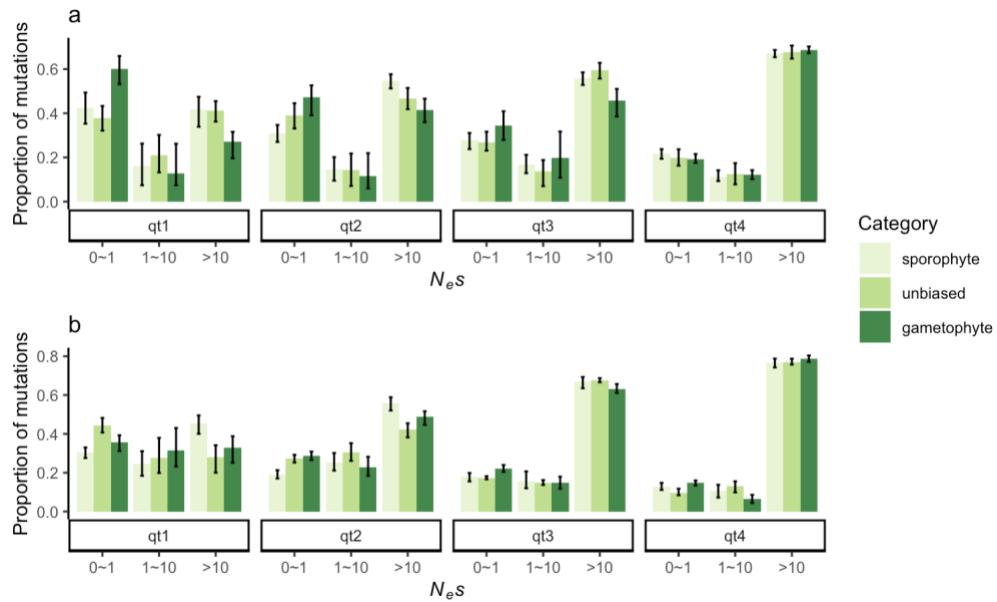


Figure B7. Distribution of fitness effects of life stage-specific genes across expression level quantiles in *R. hastatulus* (a) and *C. purpureus* (b). Number of genes in each quantile of expression level: 322, 407, 457, 525 (s) and 570, 254, 180, 707 (g) for *R. hastatulus*; 873, 429, 197, 157 (s) and 489, 454, 272, 298 (g) for *C. purpureus*. The mean and 95% confidence intervals are based on 200 bootstraps of the original dataset.
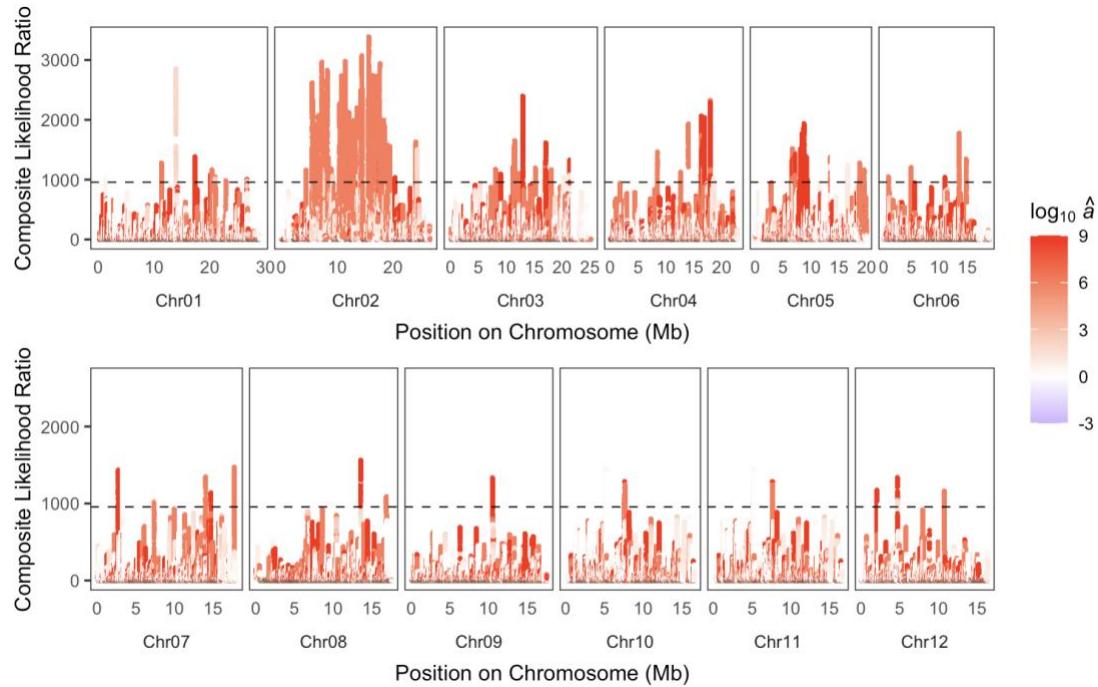
Figure B8. Genome-wide scan for balancing selection in *C. purpureus*. Black dashed lines represent the cutoff for the signal of balancing selection. $\hat{a}$: estimated dispersion parameter, a positive $\log_{10}\hat{a}$ value suggests balancing selection.
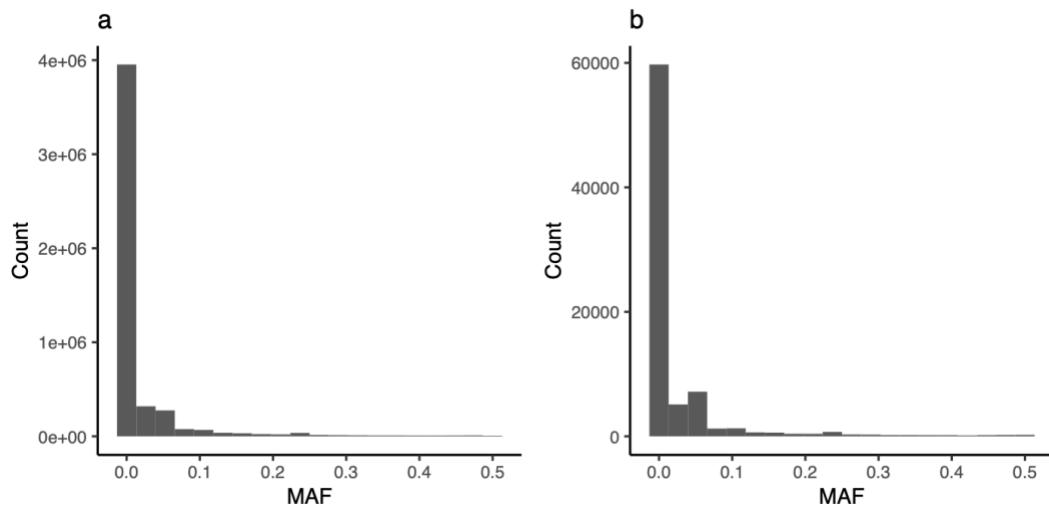


Figure B9. Distribution of minor allele frequency (MAF) of all sites tested (a) and candidate sites under balancing selection (b) in *R. hastatulus*. Mean MAF: 0.02 (a), 0.025 (b); median MAF: 0 (a), 0 (b); proportion of sites with MAF > 0.3: 0.016 (a), 0.021 (b).
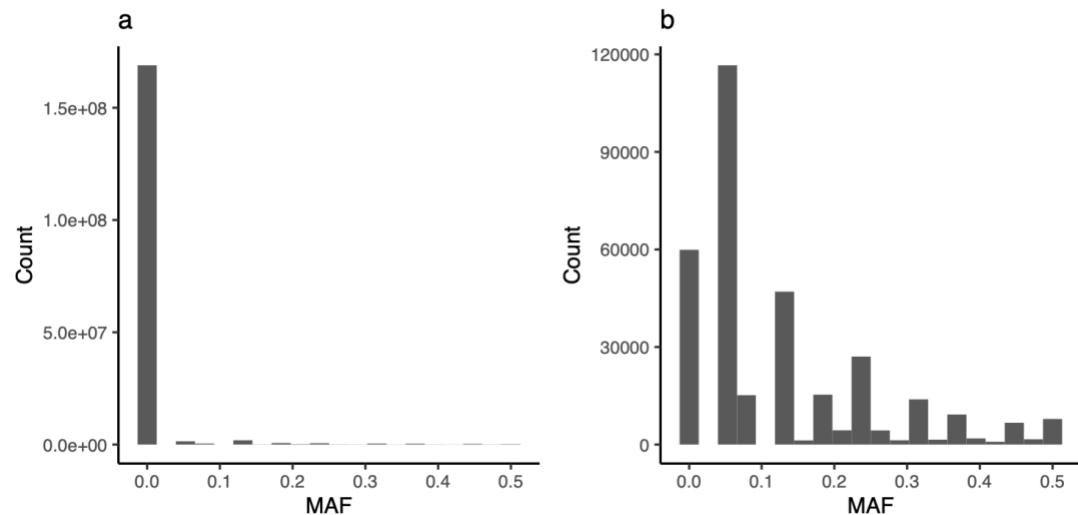
Figure B10. Distribution of minor allele frequency (MAF) of all sites tested (a) and candidate sites under balancing selection (b) in *C. purpureus*. Mean MAF: 0.0072 (a), 0.13 (b); median MAF: 0 (a), 0.063 (b); proportion of sites with MAF > 0.3: 0.008 (a), 0.13 (b).
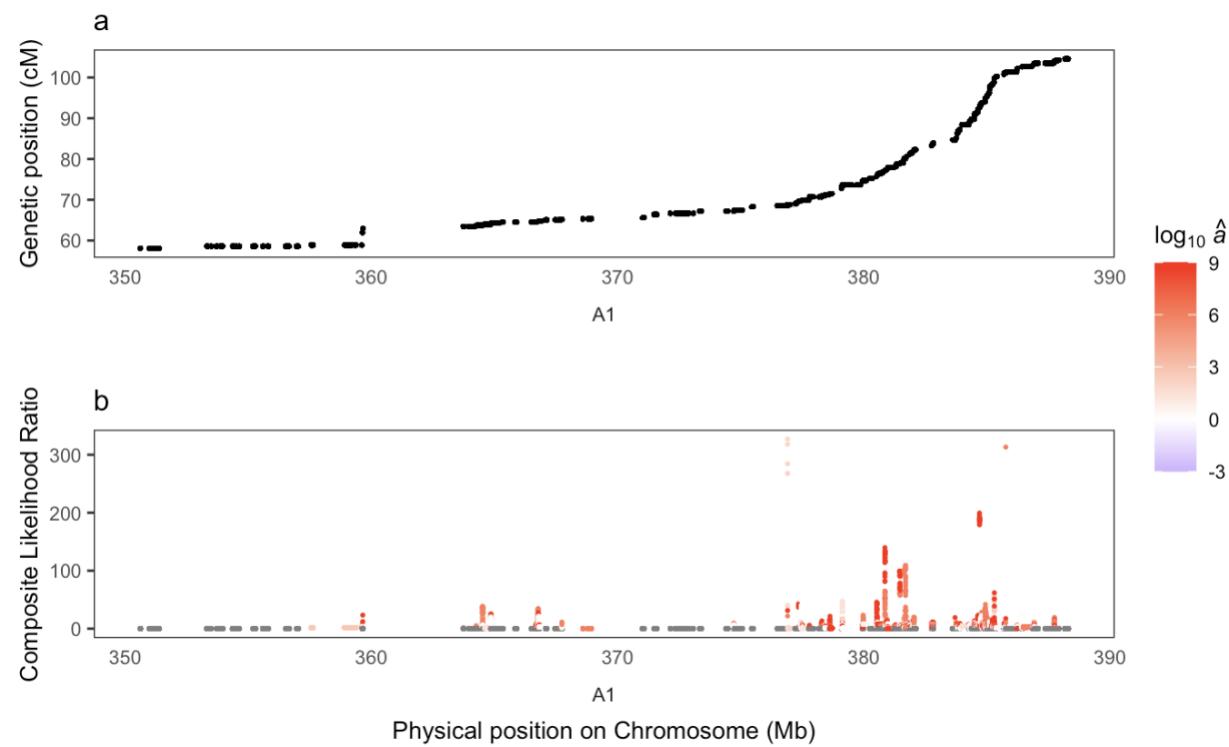


Figure B11. Scans of balancing selection in *R. hastatulus* in a 40 Mb window at the end of A1. y-axis: genetic positions of sites being tested (a), composite likelihood ratio (b).

# References

Carey SB, Peniston JH, Payton AC, Kim M, Lipzen A, Bauer D, Lail K, Daum C, Barry K, Jenkins J, et al. 2021. Novel insights into joint estimations of demography, mutation rate, and selection using UV sex chromosomes [Preprint]. bioRxiv. doi:10.1101/2021.03.30.437085.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics. 177(4):2251–2261.

Kent TV. 2023. Evolutionary consequences of plant genome structure [dissertation]. [Toronto (ON)]: University of Toronto.

Lyons E, Freeling M. 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 53(4):661–673.

Sacchi B, Humphries Z, Kružlicová J, Bodláková M, Pyne C, Choudhury BI, Gong Y, Bačovský V, Hobza R, Barrett SCH, et al. 2024. Phased assembly of neo-sex chromosomes reveals extensive y degeneration and rapid genome evolution in *Rumex hastatulus*. Mol Biol Evol. 41(4):msae074.

Song B, Marco-Sola S, Moreto M, Johnson L, Buckler ES, Stitzer MC. 2022. AnchorWave: Sensitive alignment of genomes with high sequence diversity, extensive structural polymorphism, and whole-genome duplication. Proc Natl Acad Sci. 119(1):e2113075119.

# Appendix C: Chapter 4 Supplementary Materials

## Chapter 4 Supplementary Tables

Table C1. DNA samples from *Rumex hastatulus*. SRA accessions were reported in NCBI

BioProject PRJNA983258. DNA concentration (conc.) was in ng/μL.

| Cross ID | Male ID | Female ID | Number of seeds | Male Leaf DNA conc. | Female Leaf DNA conc. | Pollen DNA conc. | Seed DNA conc. |
|---|---|---|---|---|---|---|---|
| 1 | 31c | 30a | 200 | 15.2 | 18.7 | 14 | 63.2 |
| 2 | 02f | 24a | 140 | 19.3 | 37.6 | 7 | 78.8 |
| 3 | 37h | 58d | 200 | 22.3 | 65.9 | 42.3 | 113 |
| 4 | 55e | 31h | 190 | 41.9 | 57 | 54.9 | 64.8 |
| 5 | 35h | 02g | 220 | 20.4 | 20.1 | 117 | 70.8 |
| 6 | 30f | 33e | 200 | 33.5 | 12.7 | 74.6 | 128 |
| 7 | 58b | 36b | 230 | 25 | 24.2 | 72.3 | 104 |
| 8 | 48d | 38g | 200 | 5.96 | 8.01 | 20.3 | 85.8 |
| 9 | 34b | 35g | 100 | 28.6 | 20 | 25.3 | 60.6 |
| 10 | 15c | 37b | 200 | 66.9 | 46.2 | 43.2 | 64.7 |

Table C2. Seed genotypes when the parents are homozygous for different alleles. ML: male leaf, FL: female leaf, S: seed. Genotypes: homozygous for the reference allele (0/0), homozygous for the alternative allele (1/1), heterozygous (0/1).

| | ML (0/0) | FL (1/1) | S (0/0) | S (0/1) | S (1/1) | S (0/1) % |
|---|---|---|---|---|---|---|
| **Cross1** | 16297 | 16297 | 0 | 66 | 16231 | 0.00404983 |
| **Cross2** | 21949 | 21949 | 0 | 42 | 21907 | 0.00191353 |
| **Cross3** | 5414 | 5414 | 0 | 18 | 5396 | 0.00332471 |
| **Cross4** | 1 | 1 | 0 | 0 | 1 | 0 |
| **Cross5** | 24419 | 24419 | 0 | 59 | 24360 | 0.00241615 |
| **Cross6** | 2381 | 2381 | 0 | 9 | 2372 | 0.00377992 |
| **Cross7** | 19982 | 19982 | 0 | 54 | 19928 | 0.00270243 |
| **Cross8** | 17250 | 17250 | 0 | 25 | 17225 | 0.00144928 |
| **Cross9** | 47552 | 47552 | 0 | 70 | 47482 | 0.00147207 |
| **Cross10** | 44177 | 44177 | 2 | 8665 | 35510 | 0.19614279 |
| | ML (1/1) | FL (0/0) | S (0/0) | S (0/1) | S (1/1) | S (0/1) % |
| **Cross1** | 12147 | 12147 | 12135 | 12 | 0 | 0.0009879 |
| **Cross2** | 25329 | 25329 | 25312 | 17 | 0 | 0.00067117 |
| **Cross3** | 4990 | 4990 | 4987 | 3 | 0 | 0.0006012 |
| **Cross4** | 14 | 14 | 14 | 0 | 0 | 0 |
| **Cross5** | 20941 | 20941 | 20927 | 14 | 0 | 0.00066854 |
| **Cross6** | 1785 | 1785 | 1781 | 4 | 0 | 0.0022409 |
| **Cross7** | 23671 | 23671 | 23645 | 26 | 0 | 0.00109839 |
| **Cross8** | 15222 | 15222 | 15211 | 11 | 0 | 0.00072264 |
| **Cross9** | 31980 | 31980 | 31966 | 14 | 0 | 0.00043777 |
| **Cross10** | 3981 | 3981 | 3009 | 972 | 0 | 0.24415976 |

Table C3. Number of mapped reads on each chromosome in pollen and male leaf of *Rumex hastatulus*. Pseudoautosomal regions were excluded. Y and X represent Y- and X-specific regions. The lengths of each chromosome or region in base pair are: A1 (465,013,087), A2 (329,321,726), A3 (221,760,844), A4 (179,199,150), X (220,000,000), and Y (458,837,879).

| Pollen | | | | | | |
|---|---|---|---|---|---|---|
| **Male ID** | **A1** | **A2** | **A3** | **A4** | **Y** | **X** |
| 31c | 209876871 | 141772724 | 108656853 | 90260649 | 151265444 | 67351274 |
| 02f | 245947219 | 169048431 | 130658397 | 106205600 | 182778166 | 78273781 |
| 37h | 211176511 | 143838311 | 108405812 | 82073812 | 155307732 | 60848251 |
| 55e | 261574754 | 172535336 | 132379557 | 108292308 | 193590015 | 76044420 |
| 35h | 215211146 | 146149262 | 113793706 | 86303163 | 165151756 | 67874416 |
| 30f | 228709917 | 160461922 | 120888768 | 96926442 | 167261623 | 67740472 |
| 58b | 271363494 | 183788677 | 142885702 | 115515118 | 199378618 | 79399641 |
| 48d | 211677055 | 146100354 | 107883745 | 88395902 | 154985179 | 61773030 |
| 15c | 219528268 | 158544042 | 95005787 | 82379579 | 137854940 | 71789116 |
| 34b | 250527284 | 175772455 | 127704450 | 91084639 | 187906934 | 80561511 |
| **Leaf** | | | | | | |
| **Male ID** | **A1** | **A2** | **A3** | **A4** | **Y** | **X** |
| 31c | 45289250 | 29931277 | 23512298 | 19456842 | 35186949 | 14344354 |
| 02f | 59818338 | 39684793 | 31478954 | 25658485 | 49416620 | 18680500 |
| 37h | 47907078 | 32047179 | 24624379 | 18535707 | 36591548 | 13783310 |
| 55e | 73982940 | 48559236 | 37423823 | 30827696 | 55900171 | 21512461 |
| 35h | 51673304 | 34227456 | 27226295 | 20370263 | 40451776 | 16480279 |
| 30f | 57426255 | 39634693 | 30309892 | 24197090 | 37293379 | 19118484 |
| 58b | 66635117 | 43901053 | 34933213 | 28007076 | 51835409 | 19258165 |
| 48d | 54622007 | 35930122 | 27540564 | 22709327 | 45041951 | 15672613 |
| 15c | 59535112 | 43511328 | 26118774 | 22324680 | 33556997 | 20891092 |
| 34b | 54147605 | 36635767 | 27185649 | 19013384 | 42092875 | 17634500 |

Table C4. Odds ratio and *p*-values from contingency tests on the number of mapped reads in leaf and pollen. OR: odds ratio, p: *p*-value. X, Y, A1, A2, A3, A4 represent different chromosomes shown in Table C3.

| Odds ratio | | | | | | |
|------|-------|-------|-------|-------|-------|-------|
| Male | Y/A1 | Y/A2 | Y/A3 | Y/A4 | Y/A | X/Y |
| 31c | 0.928 | 0.908 | 0.930 | 0.927 | 0.923 | 1.092 |
| 02f | 0.900 | 0.868 | 0.891 | 0.894 | 0.889 | 1.133 |
| 37h | 0.963 | 0.946 | 0.964 | 0.959 | 0.958 | 1.040 |
| 55e | 0.980 | 0.975 | 0.979 | 0.986 | 0.979 | 1.021 |
| 35h | 0.980 | 0.956 | 0.977 | 0.964 | 0.971 | 1.009 |
| 30f | 1.126 | 1.108 | 1.125 | 1.120 | 1.120 | 0.790 |
| 58b | 0.945 | 0.919 | 0.940 | 0.933 | 0.935 | 1.072 |
| 48d | 0.888 | 0.846 | 0.878 | 0.884 | 0.874 | 1.145 |
| 34b | 0.965 | 0.930 | 0.950 | 0.932 | 0.948 | 1.023 |
| 15c | 1.114 | 1.127 | 1.129 | 1.113 | 1.120 | 0.836 |
| P-value | | | | | | |
| Male | Y/A1 | Y/A2 | Y/A3 | Y/A4 | Y/A | X/Y |
| 31c | 0 | 0 | 0 | 0 | 0 | 0 |
| 02f | 0 | 0 | 0 | 0 | 0 | 0 |
| 37h | 0 | 0 | 0 | 0 | 0 | 0 |
| 55e | 0 | 0 | 0 | 0 | 0 | 0 |
| 35h | 0 | 0 | 0 | 0 | 0 | 1.77E-158 |
| 30f | 0 | 0 | 0 | 0 | 0 | 0 |
| 58b | 0 | 0 | 0 | 0 | 0 | 0 |
| 48d | 0 | 0 | 0 | 0 | 0 | 0 |
| 34b | 0 | 0 | 0 | 0 | 0 | 0 |
| 15c | 0 | 0 | 0 | 0 | 0 | 0 |
| Odds ratio | | | | | | |
| Male | X/A1 | X/A2 | X/A3 | X/A4 | X/A | |
| 31c | 1.013 | 0.991 | 1.016 | 1.012 | 1.008 | |
| 02f | 1.019 | 0.984 | 1.010 | 1.012 | 1.007 | |
| 37h | 1.001 | 0.984 | 1.003 | 0.997 | 0.996 | |
| 55e | 1.000 | 0.995 | 0.999 | 1.006 | 0.999 | |
| 35h | 0.989 | 0.965 | 0.985 | 0.972 | 0.979 | |
| 30f | 0.890 | 0.875 | 0.888 | 0.885 | 0.885 | |
| 58b | 1.012 | 0.985 | 1.008 | 1.000 | 1.002 | |
| 48d | 1.017 | 0.969 | 1.006 | 1.013 | 1.002 | |
| 34b | 0.987 | 0.952 | 0.973 | 0.954 | 0.970 | |
| 15c | 0.932 | 0.943 | 0.945 | 0.931 | 0.937 | |

| P-value | | | | | | |
|---|---|---|---|---|---|---|
| **Male** | **X/A1** | **X/A2** | **X/A3** | **X/A4** | **X/A** | |
| **31c** | 0 | 2.25E-135 | 0 | 3.27E-217 | 2.15E-145 | |
| **02f** | 0 | 0 | 2.66E-186 | 1.47E-285 | 8.82E-140 | |
| **37h** | 1.02E-05 | 0 | 8.58E-14 | 2.84E-14 | 3.34E-31 | |
| **55e** | 4.70E-01 | 1.47E-68 | 2.64E-02 | 3.05E-86 | 4.70E-02 | |
| **35h** | 1.40E-275 | 0 | 0.00E+00 | 0 | 0 | |
| **30f** | 0 | 0 | 0 | 0 | 0 | |
| **58b** | 0 | 0 | 2.50E-139 | 0.242695412 | 2.05E-18 | |
| **48d** | 0 | 0 | 1.23E-67 | 3.87E-253 | 3.85E-08 | |
| **34b** | 0 | 0 | 0 | 0 | 0 | |
| **15c** | 0 | 0 | 0 | 0 | 0 | |

# Chapter 4 Supplementary Figures
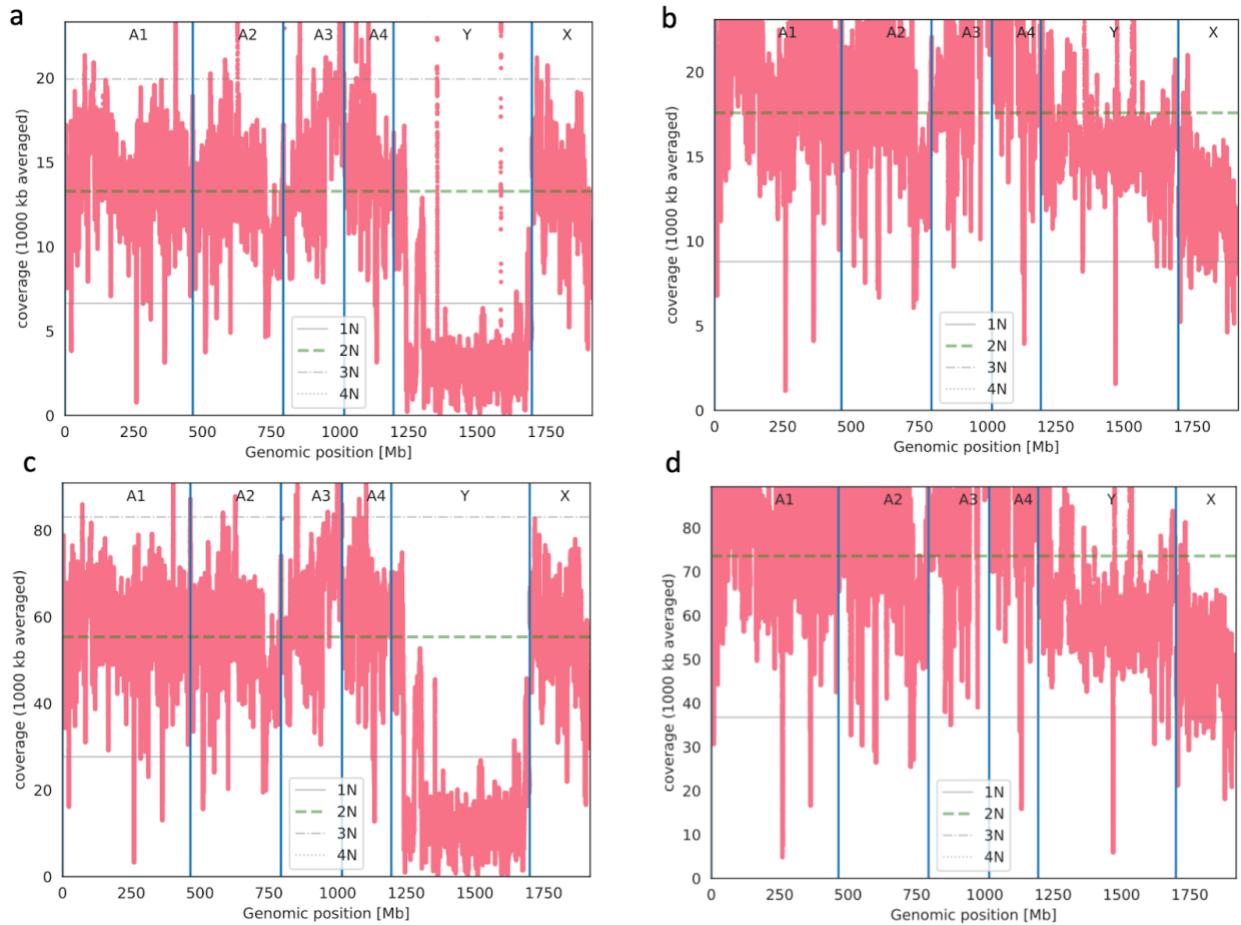


Figure C1. Coverage across the genome in female leaf (a), male leaf (b), seed (c), and pollen (d) in cross 7 (58b × 36b). The first 45 Mb of Y chromosome is the pseudoautosomal region.
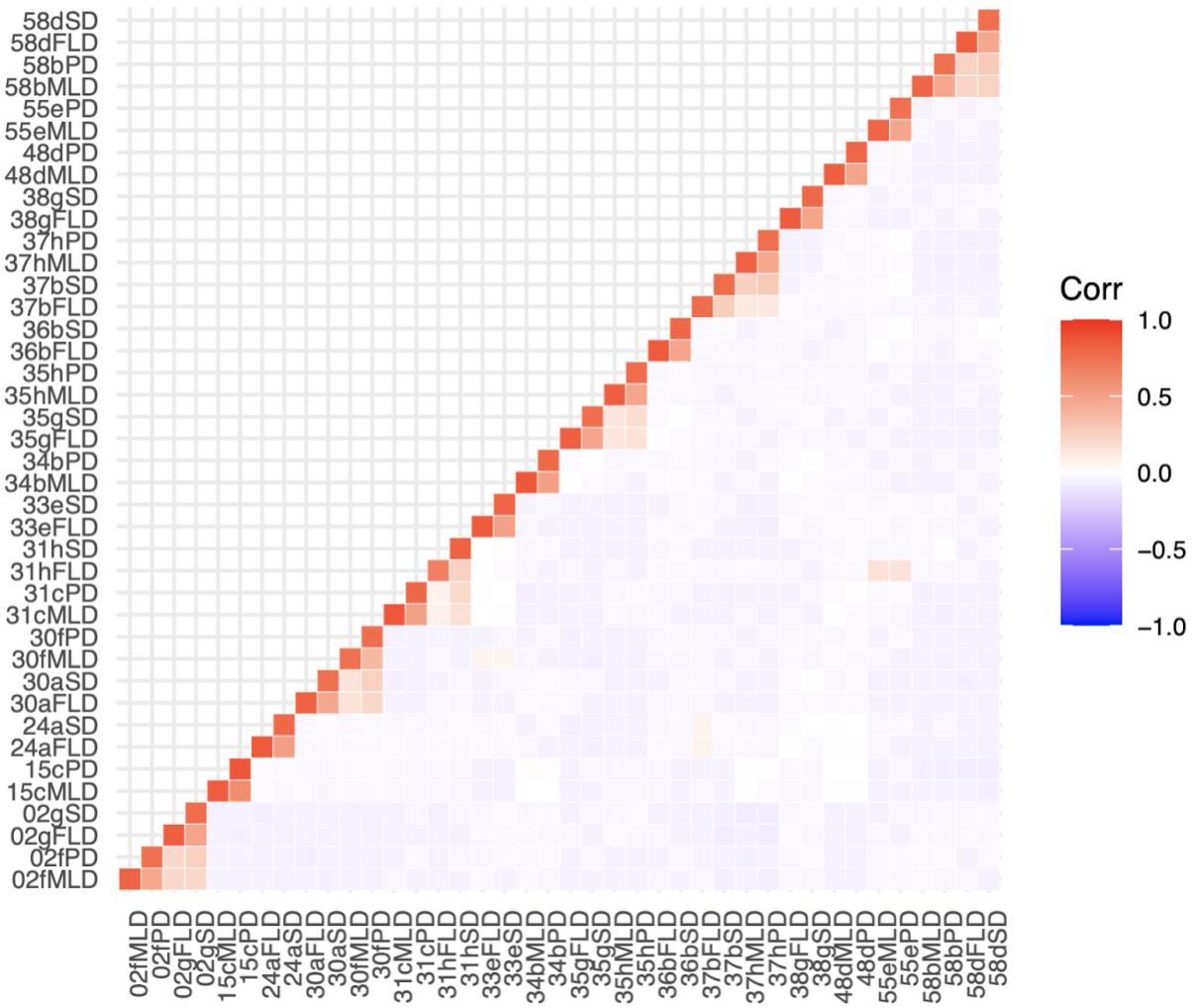
Figure C2. Genetic relatedness matrix of all 40 DNA samples. FLD: female leaf DNA, MLD: male leaf DNA, SD: seed DNA, PD: pollen DNA. Cross pairing is available in Supplementary Table C1.
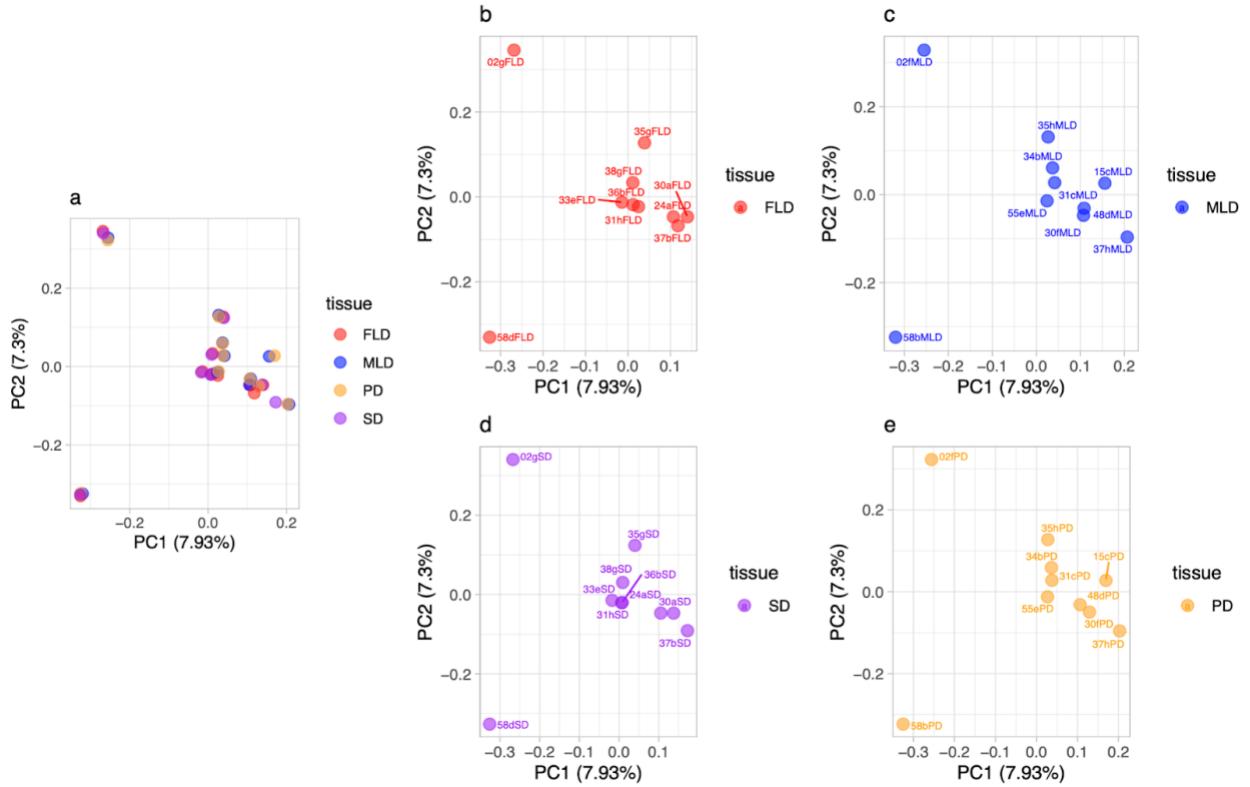
Figure C3. PCA of DNA samples used in this study. a. all samples, b. female leaf, c. male leaf, d. seed, e. pollen. Cross pairing is available in Supplementary Table C1.
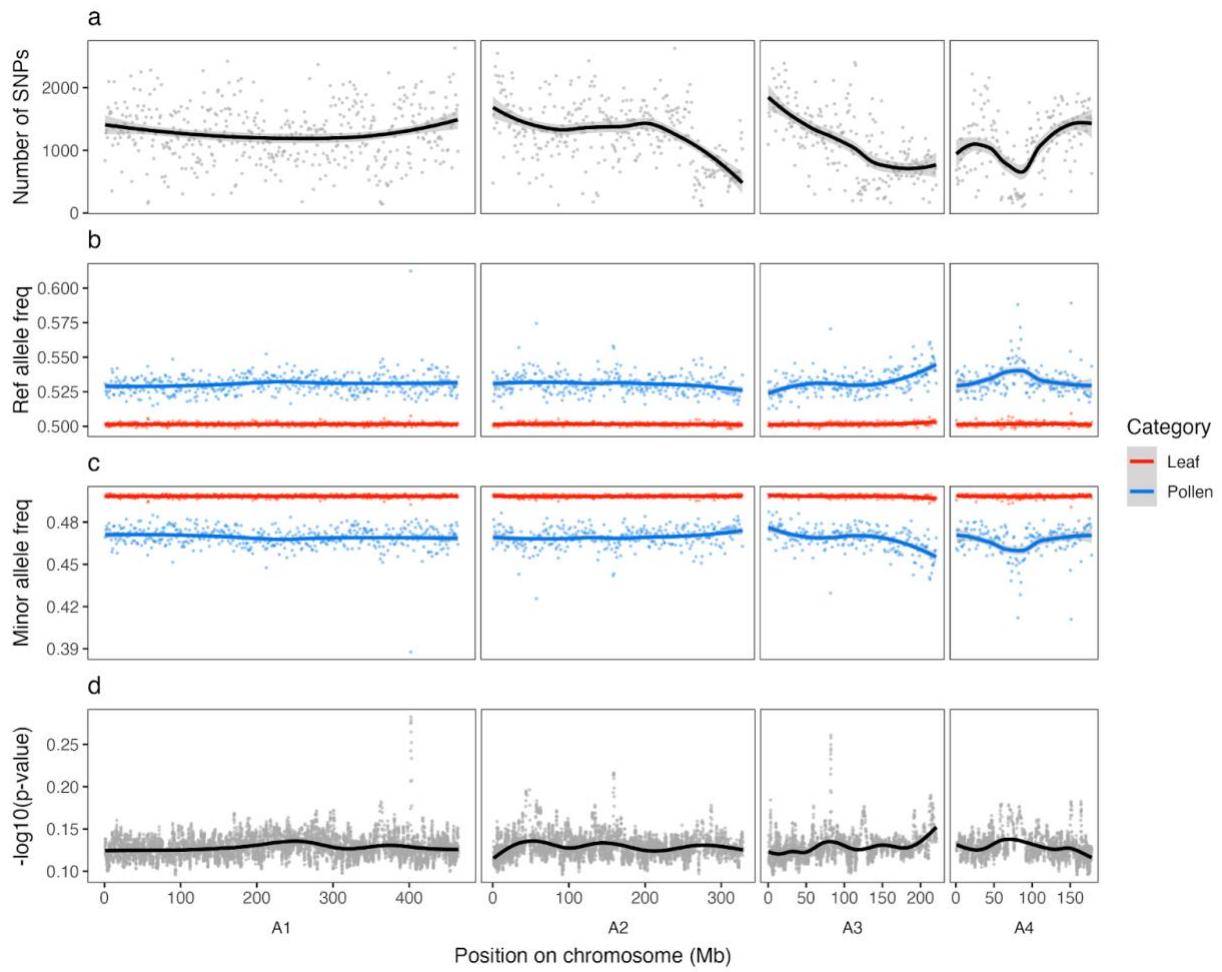
Figure C4. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 1. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
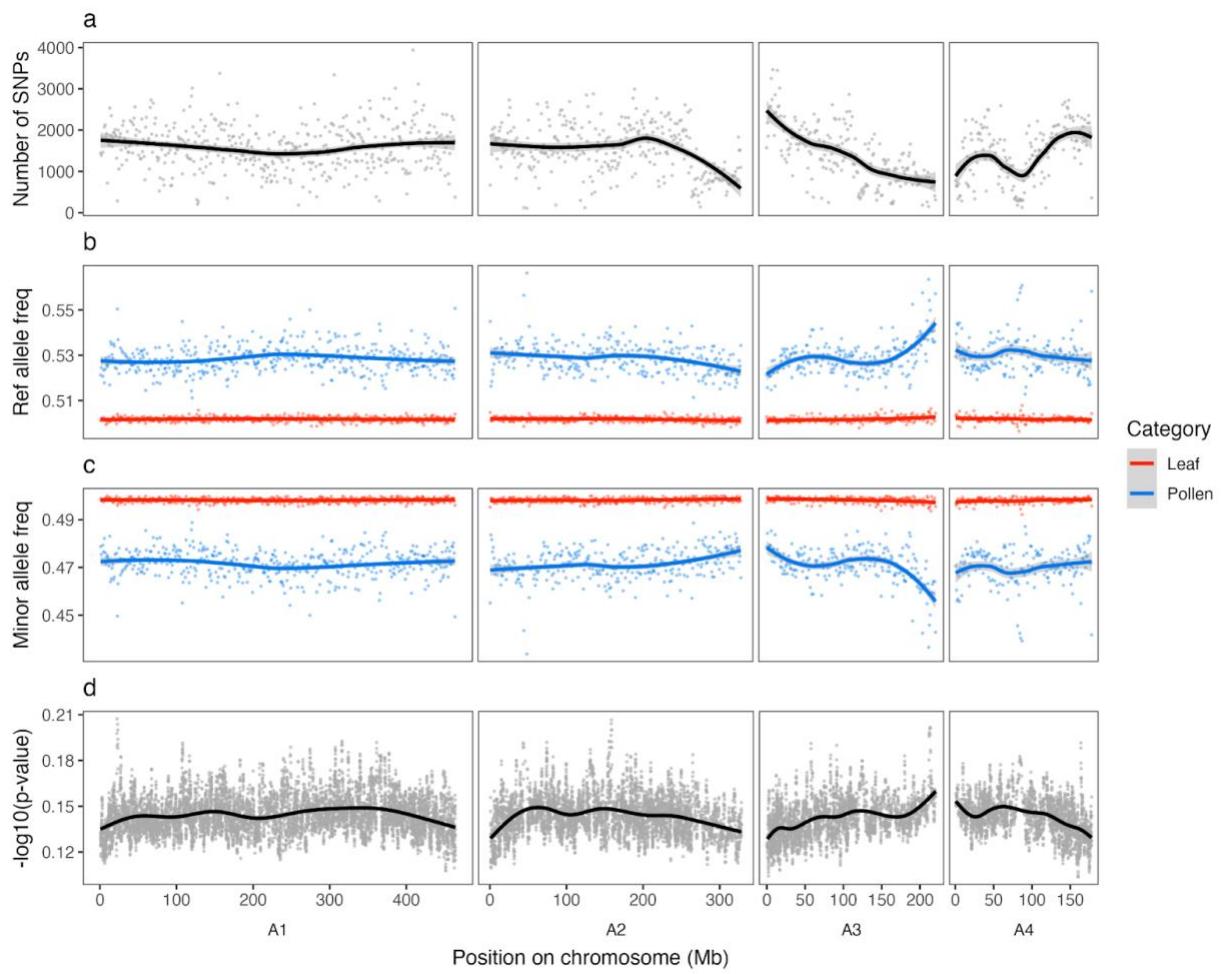
Figure C5. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 2. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
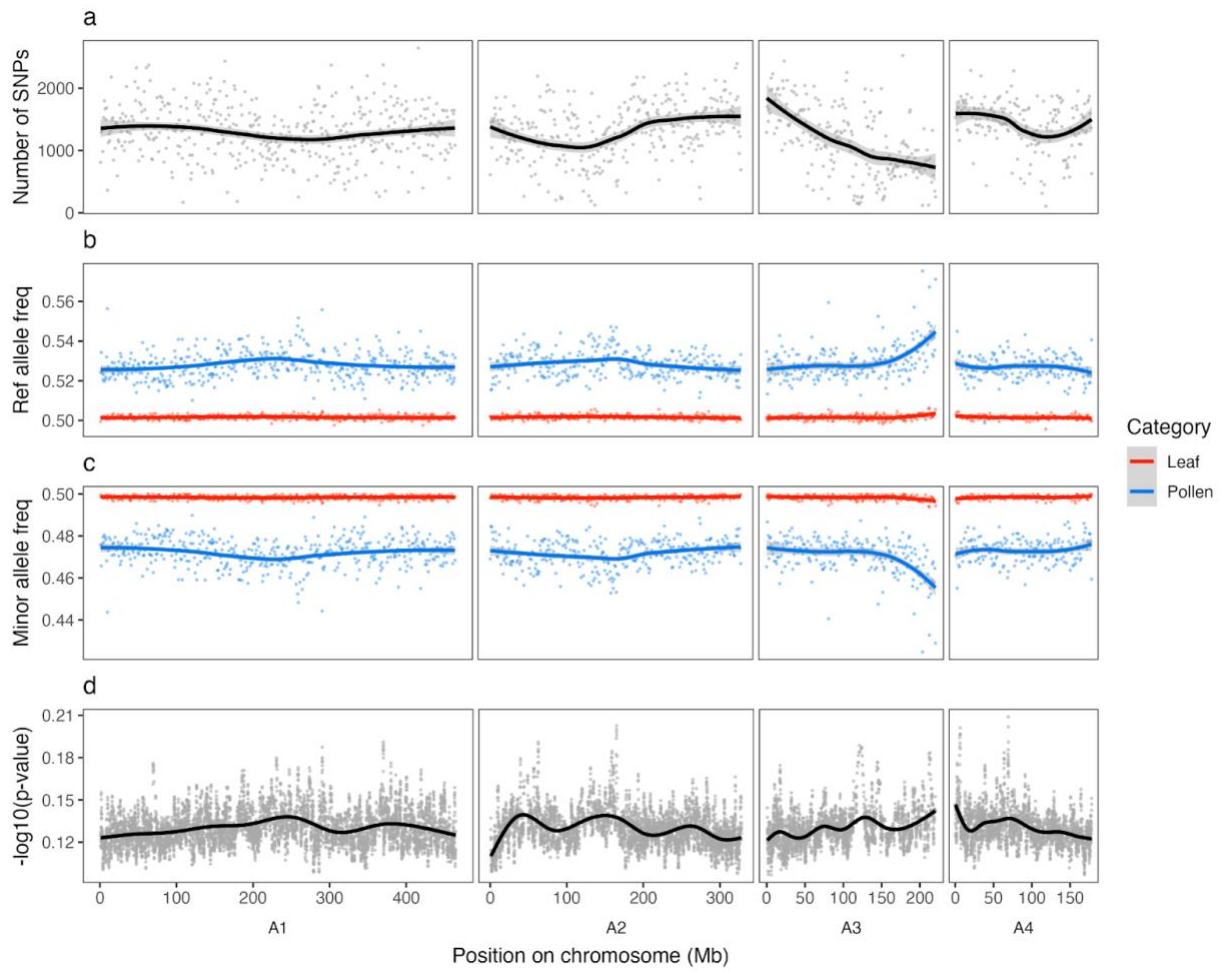
Figure C6. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 3. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
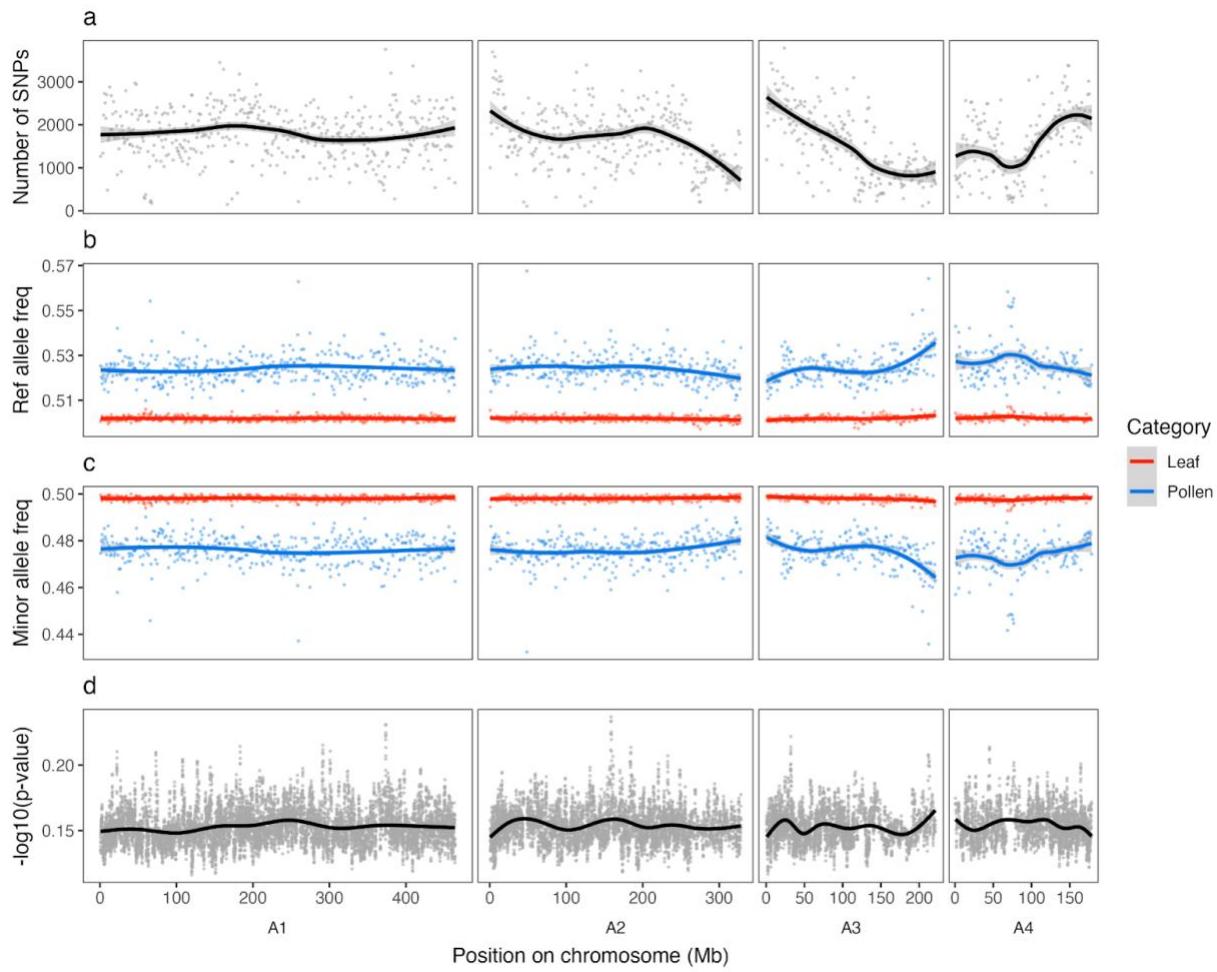
Figure C7. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 4. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
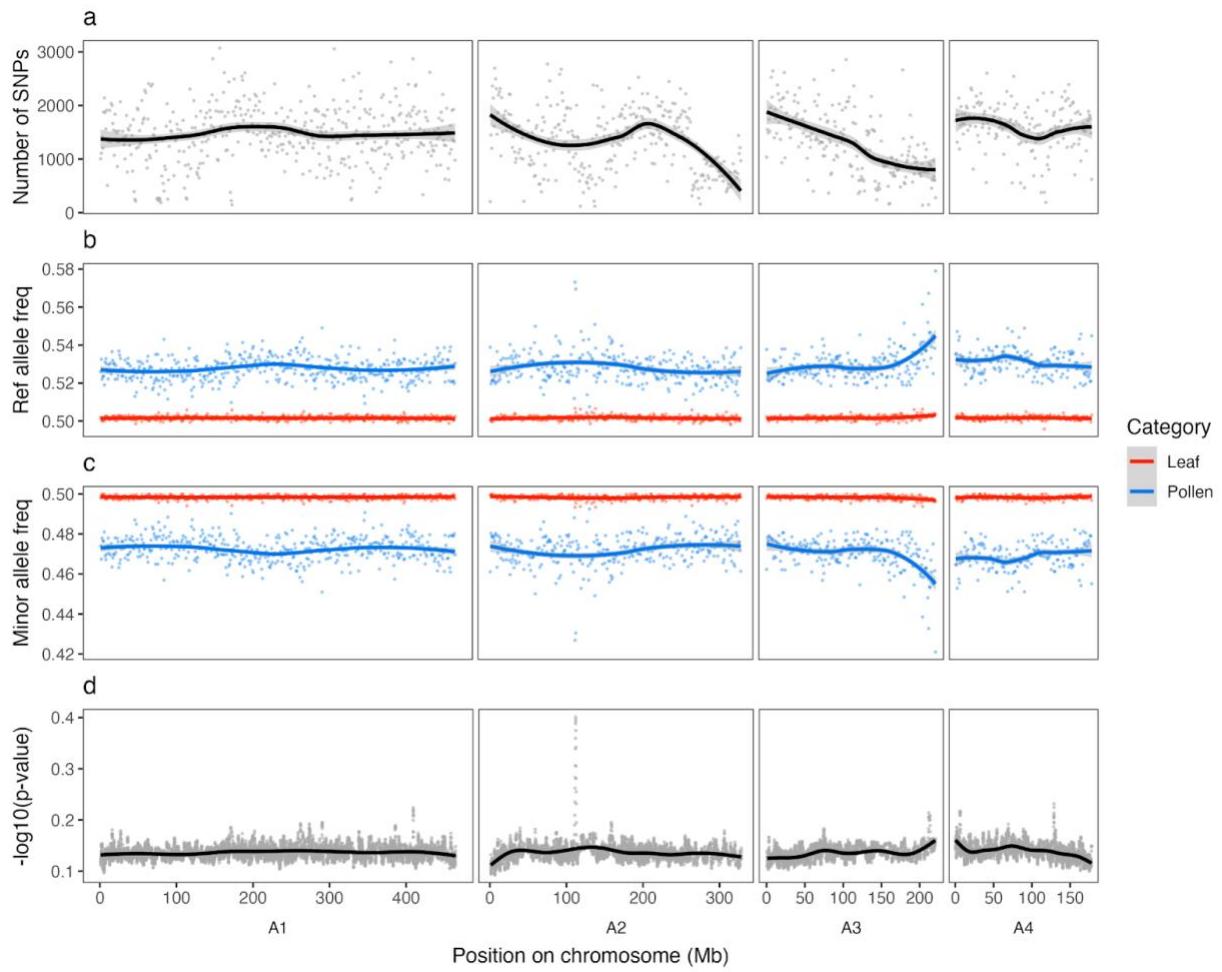
Figure C8. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 5. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
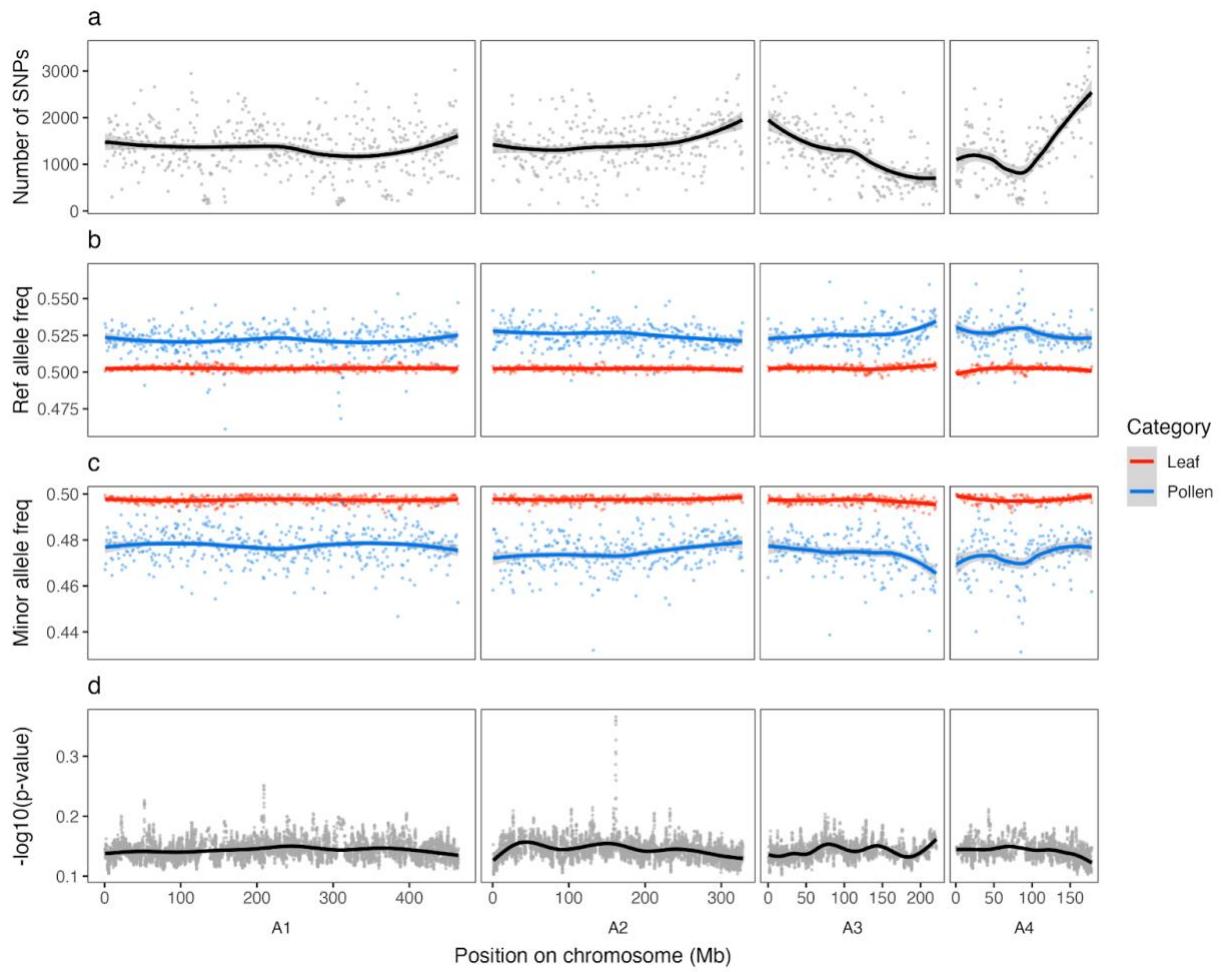
Figure C9. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 6. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
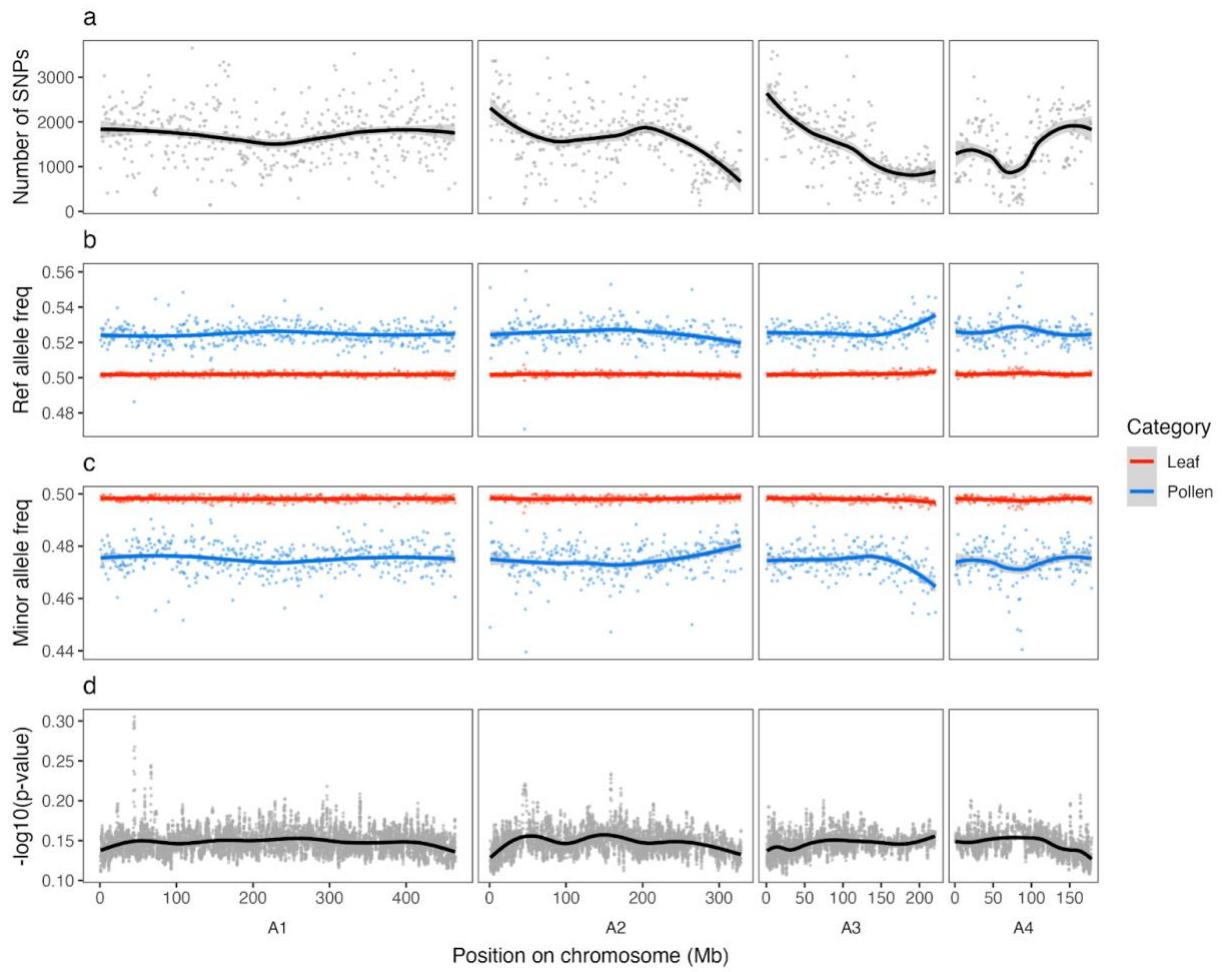
Figure C10. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 7. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
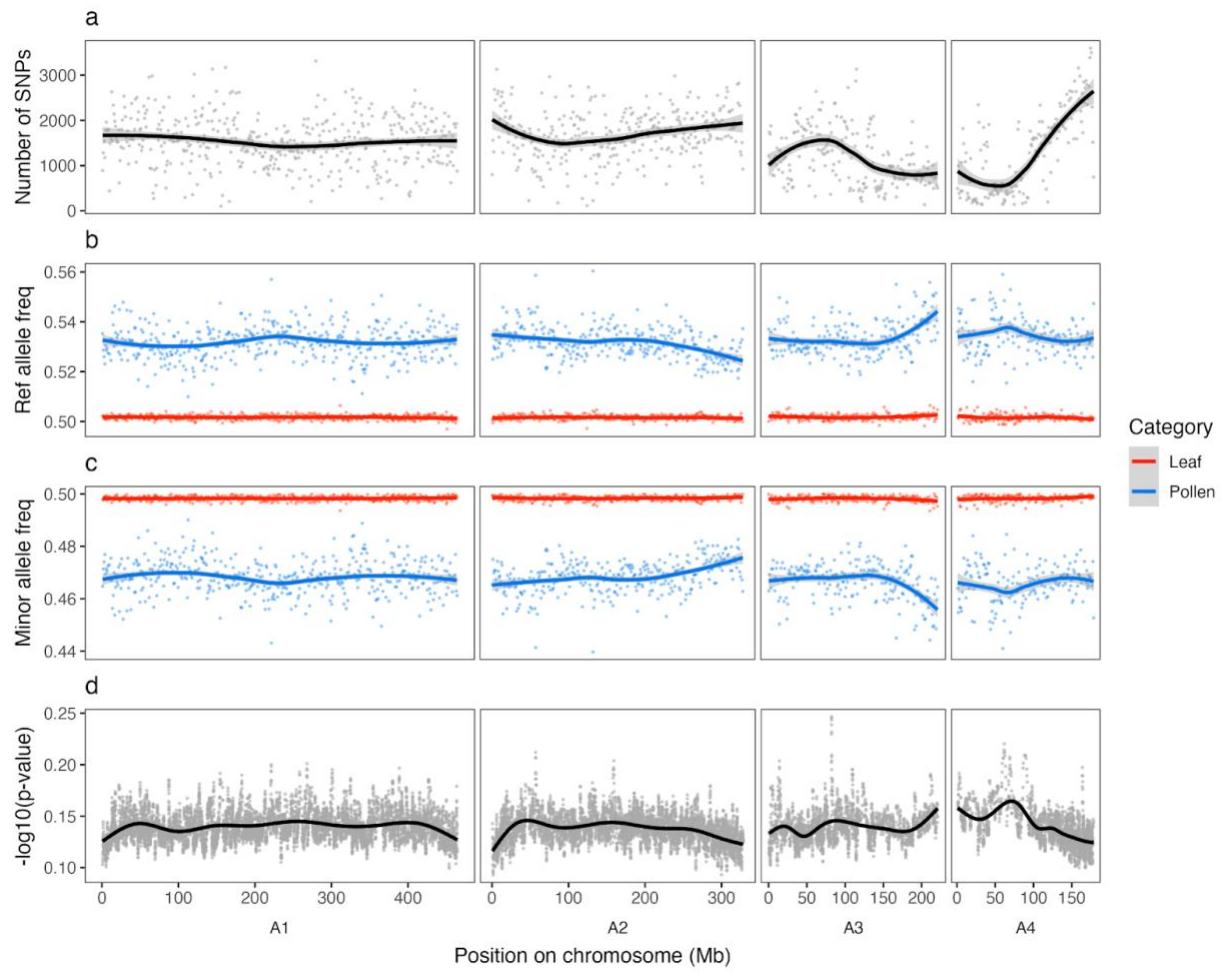
Figure C11. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 9. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.
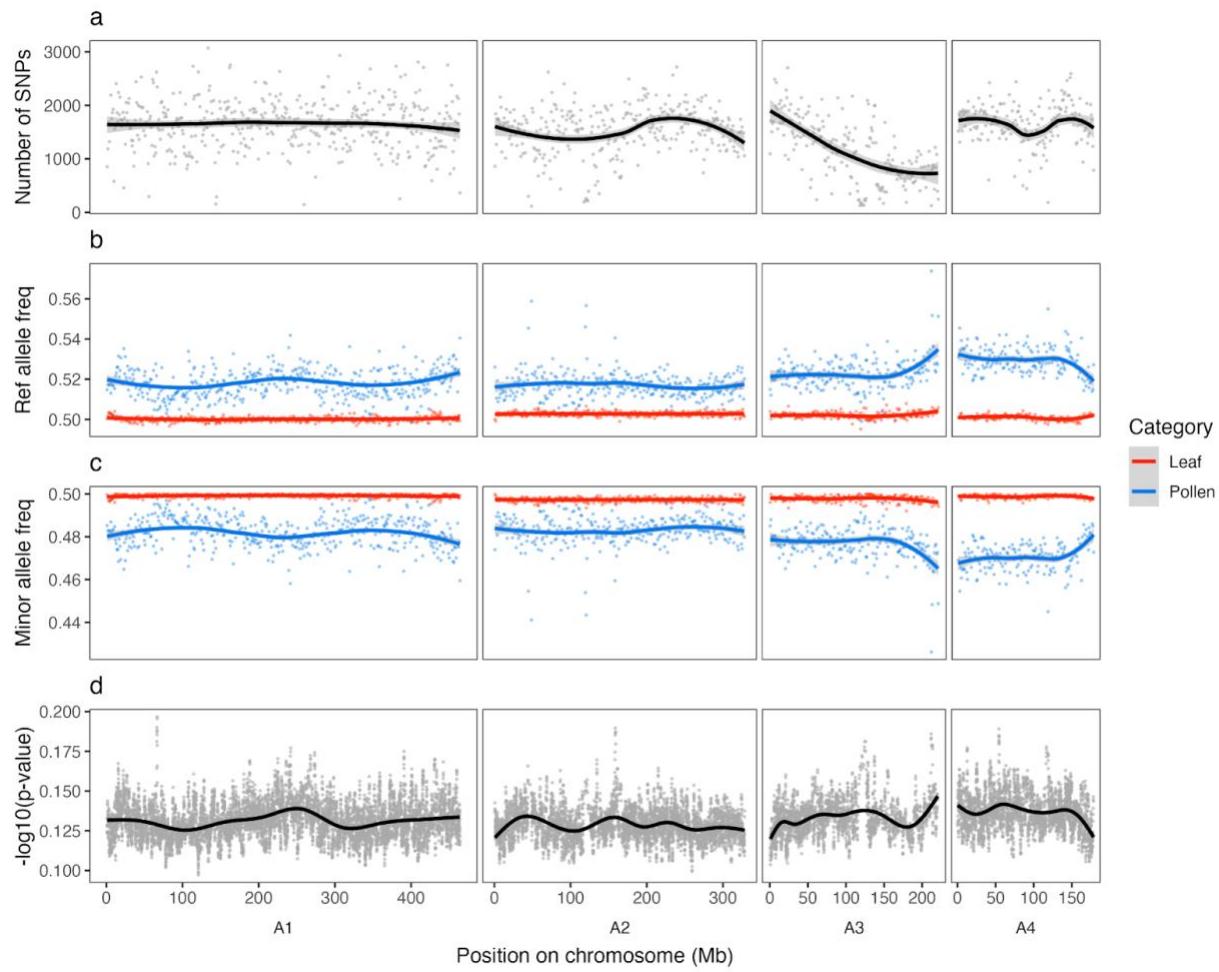
Figure C12. Allele frequencies at heterozygous SNPs in male leaf and pollen of Male 10. a. number of heterozygous SNPs, b. reference allele frequency, c. minor allele frequency, d. average *p*-values from Fisher's exact tests on allelic depths in leaf and pollen. a-c: window size = 1 Mb, step size = 1 Mb; d: window size = 1000 SNPs, step size = 100 SNPs. Smoothed lines were generated by the default smoothing function in R.