











# Testing for the Genomic Footprint of Conflict Between Life Stages in an Angiosperm and Moss Species

Meng Yuan <sup>1,\*</sup>, Leslie M. Kollar <sup>2,3</sup>, Bianca M. Sacchi <sup>1</sup>, Sarah B. Carey <sup>2,4</sup>, Baharul I. Choudhury <sup>1,5</sup>, Teresa Jones<sup>4,6</sup>, Jane Grimwood <sup>4,6</sup>, Spencer C.H. Barrett <sup>1</sup>, Stuart F. McDaniel <sup>2,†</sup>, Stephen I. Wright <sup>1,†</sup>, John R. Stinchcombe <sup>1,†</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

<sup>2</sup>Department of Biology, University of Florida, Gainesville, FL, USA

<sup>3</sup>Office of Research Development, University of Michigan, Ann Arbor, MI, USA

<sup>4</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

<sup>5</sup>Department of Biology, Queen's University, Kingston, ON, Canada

<sup>6</sup>HudsonAlpha Genome Sequencing Center, Huntsville, AL, USA

<sup>†</sup>S.F.M., S.I.W., and J.R.S. contributed equally.

\*Corresponding author: E-mail: my.yuan@mail.utoronto.ca.

Accepted: July 08, 2025

## 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.

**Key words:** balancing selection, antagonistic pleiotropy, intralocus conflict, gene expression, gametophytic selection.

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

## Significance

The genome-wide importance of balancing selection has been a key question in evolutionary biology. Despite theoretical support for the potential of balancing selection generated by antagonistic pleiotropy between plant life stages, to our knowledge, no empirical studies have systematically investigated this on a genome-wide scale. Our results revealed genome-wide patterns more consistent with synergistic pleiotropy, with gametophyte-biased genes more likely to be under relaxed purifying selection in both the angiosperm and moss species we investigated. In addition, our study suggests gene expression level and breadth has a greater effect on selection efficacy than differences between ploidy levels in different life stages.

## 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 versus 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. 2022; 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 the 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 Ruzicka 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 *C. 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 suggest

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 (Table S1). 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, Table S3, see Materials and Methods). In both species, there is a large overlap in gene expression between the life stages (Table 1). In *R. hastatulus*, we identified 15,958 gametophyte-expressed genes, consisting of 52% of all

**Table 1** Summary of samples used and gene expression analyses in *Rumex hastatulus* and *Ceratodon 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

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

All percentages are based on comparison to the number of expressed genes.

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.

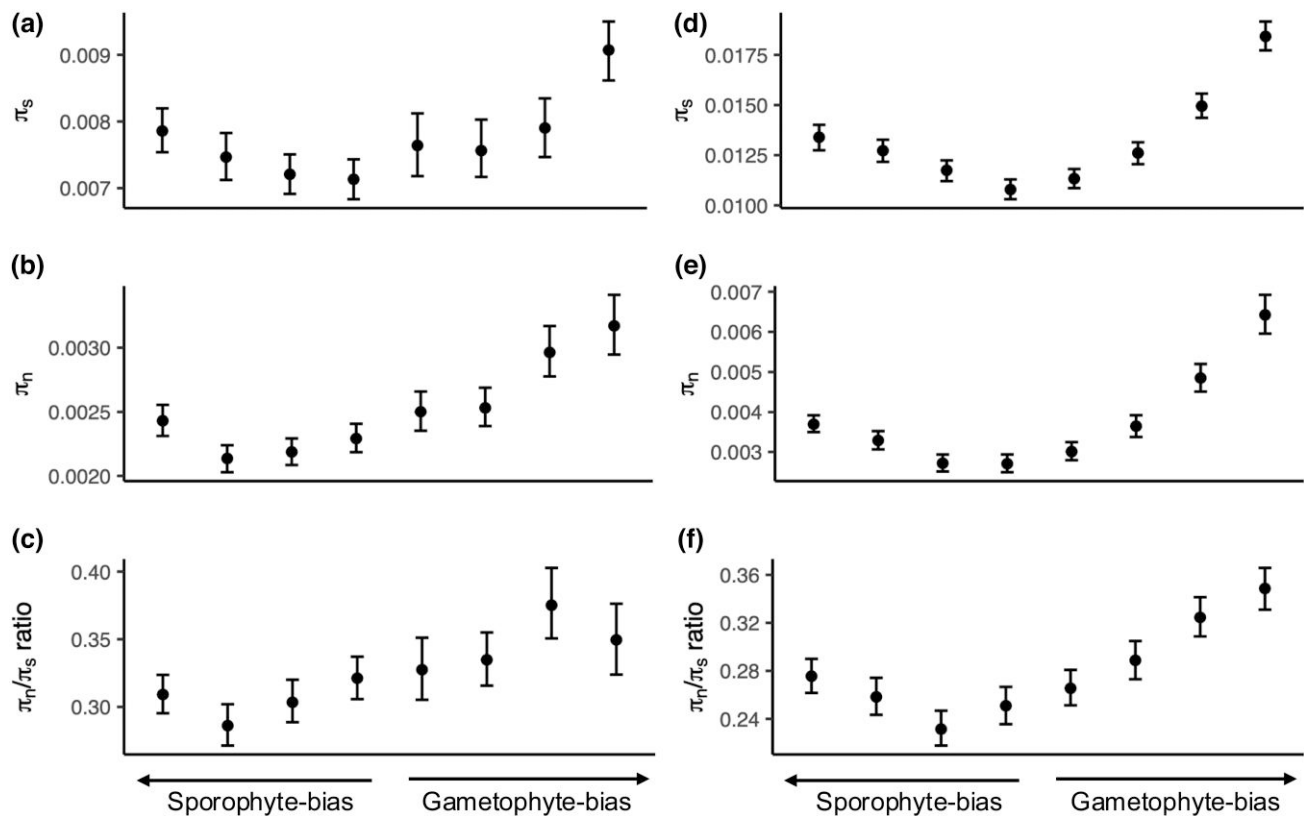
We first identified genes that were differentially expressed (DE) between life stages and exclusively expressed in one life stage (Table S4). There was a slightly higher percentage of overlapping gene expression and a lower percentage of differentially expressed genes between life stages in *C. purpureus* than *R. hastatulus* (Table 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 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 (Tables S5 and S6), 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 Fila 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 (Figs. S1 and S2). 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 (Fig. 1, Table S7). 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 the greatest for the highly gametophytic-biased genes (Fig. 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 (Fig. 1a and b). 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 (Fig. S3).

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 (Fig. 1c and f), suggesting a reduction in the strength of purifying selection on the most gametophyte-biased genes.



**Fig. 1.** Effect of expression bias between life stages on nucleotide diversity in *Rumex hastatulus* (a–c) and *Ceratodon purpureus* (d–f). x axis: quantiles of log2FoldChange between gametophyte and sporophyte expression. Number of genes in each bin: 1,009 to 1,847 for *R. hastatulus*; 1,943 to 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.

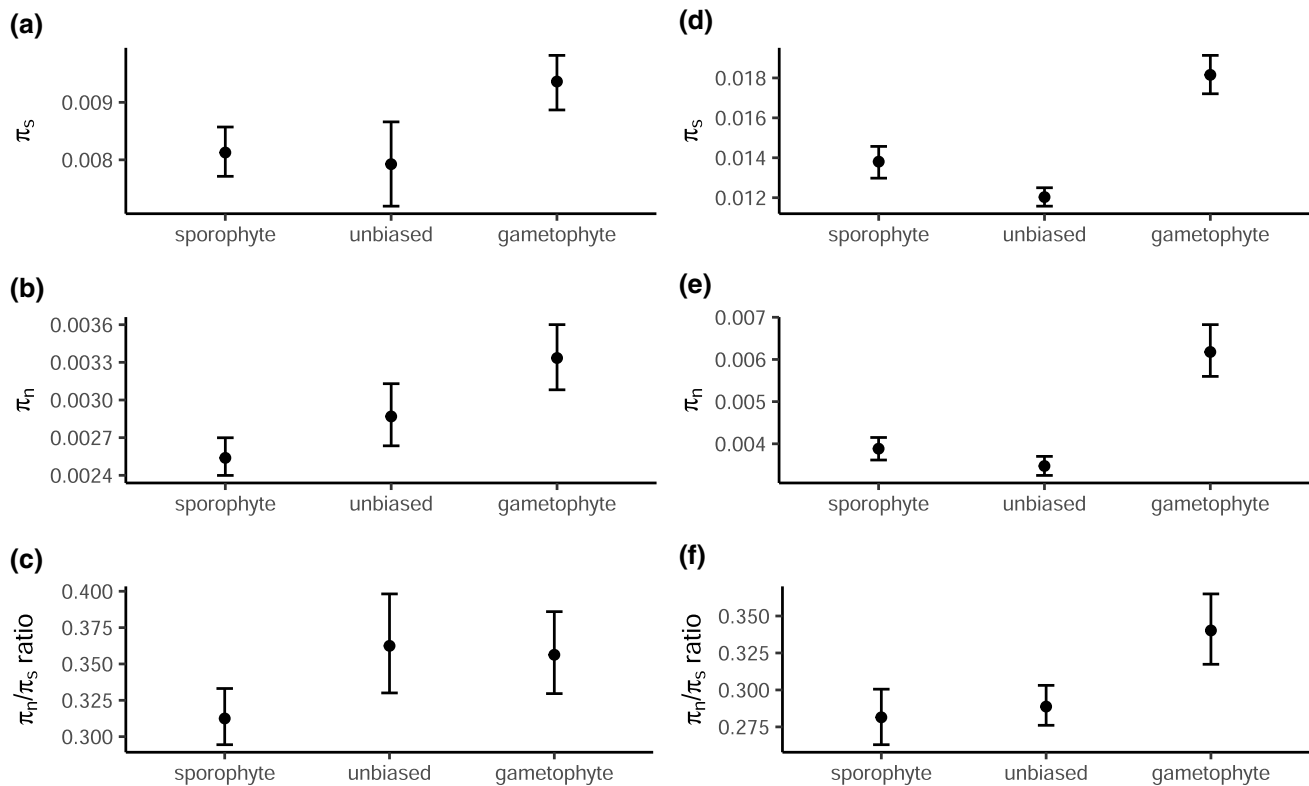
To further distinguish between the two alternative explanations for increased nucleotide diversity in gametophyte biased genes, we examined diversity patterns at genes specifically expressed in one of the two life stages, since these genes should experience the least pleiotropy across stages with reduced expression breadth. We compared  $\pi_s$ ,  $\pi_n$ , and  $\pi_n/\pi_s$  ratio for gametophyte-specific, sporophyte-specific and unbiased genes (Fig. 2, Fig. S4; Table S8). The results are mostly consistent with Fig. 1, with gametophyte-specific genes having a higher level of  $\pi_s$  and  $\pi_n$  than sporophyte-specific and unbiased genes in both species (Fig. 2a, b, d, and e), although the  $\pi_n/\pi_s$  ratios were not elevated above unbiased genes in *R. hastatulus* as in *C. purpureus* (Fig. 2c and f).

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  $\pi_s$  (Fig. 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* (Fig. 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.

#### 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



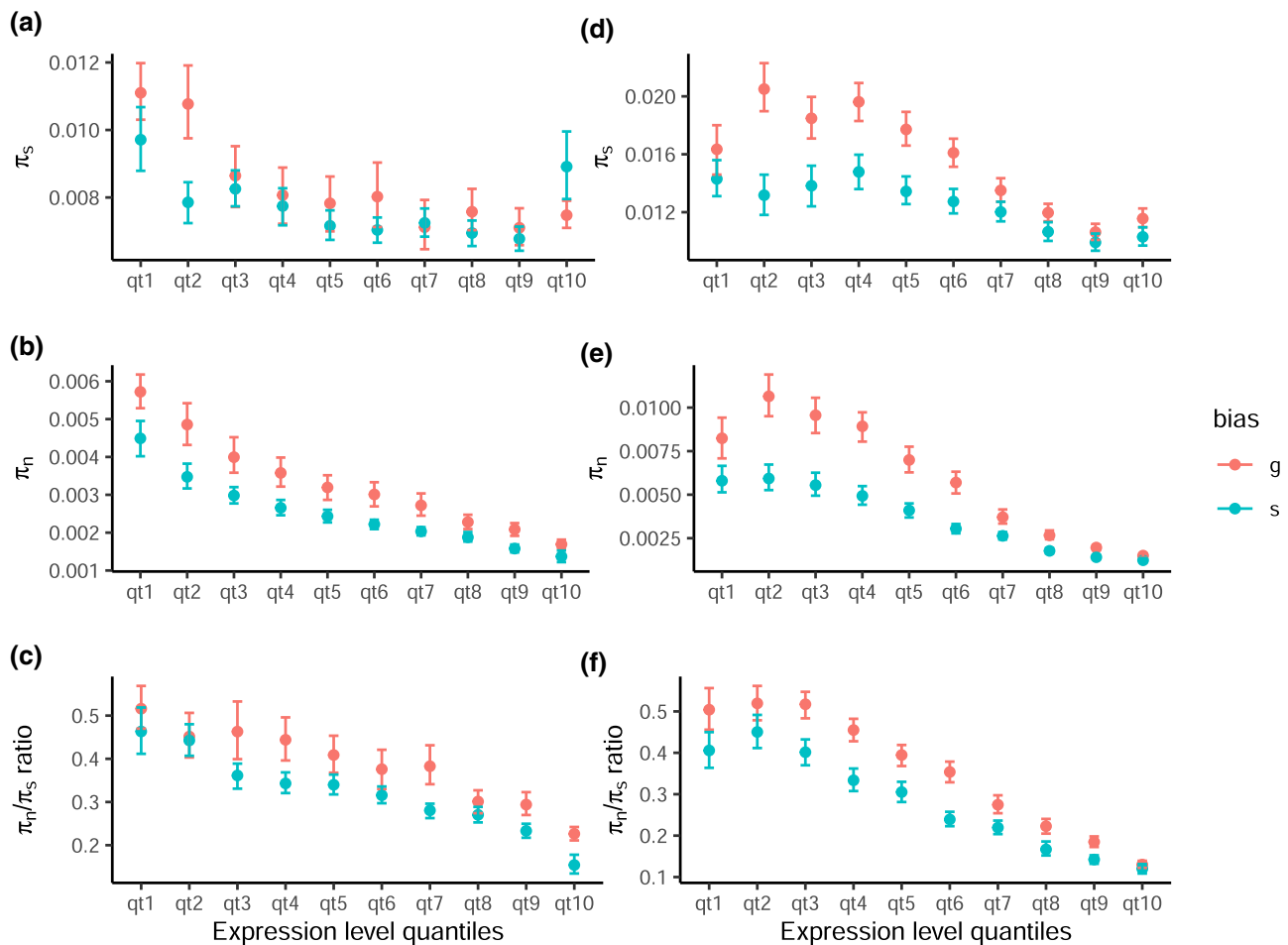
**Fig. 2.** Weighted mean nucleotide diversity of gametophyte-specific, sporophyte-specific, unbiased genes in *Rumex hastatulus* (a–c) and *Ceratodon purpureus* (d–f). Number of genes in each category: 1,149 (s), 920 (g), 524 (unbiased) for *R. hastatulus*; 1,232 (s), 1,217 (g), 3,380 (unbiased) for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 1,000 bootstraps of the original dataset. Note that the scales on the y axis are different between species.

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 (Fig. 4). In both species,  $D_n$  was smaller than  $D_s$ , consistent with purifying selection on nonsynonymous sites (Fig. 4a versus b, c versus d). 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 (Fig. 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 (Fig. 4c and d). 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 (Fig. S5). Gametophyte- and sporophyte-biased genes had similar levels of  $D_s$  and  $D_n$  across expression levels in both species. Combined with patterns of  $\pi$  (Figs. 1 to 3), our results are consistent with the hypothesis that gametophyte-biased genes are under relaxed purifying selection.

#### 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 (Fig. S6, Table S9), then we binned the gene sets into four expression level quantiles (Fig. S7, Table S10). 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 (Fig. S7).

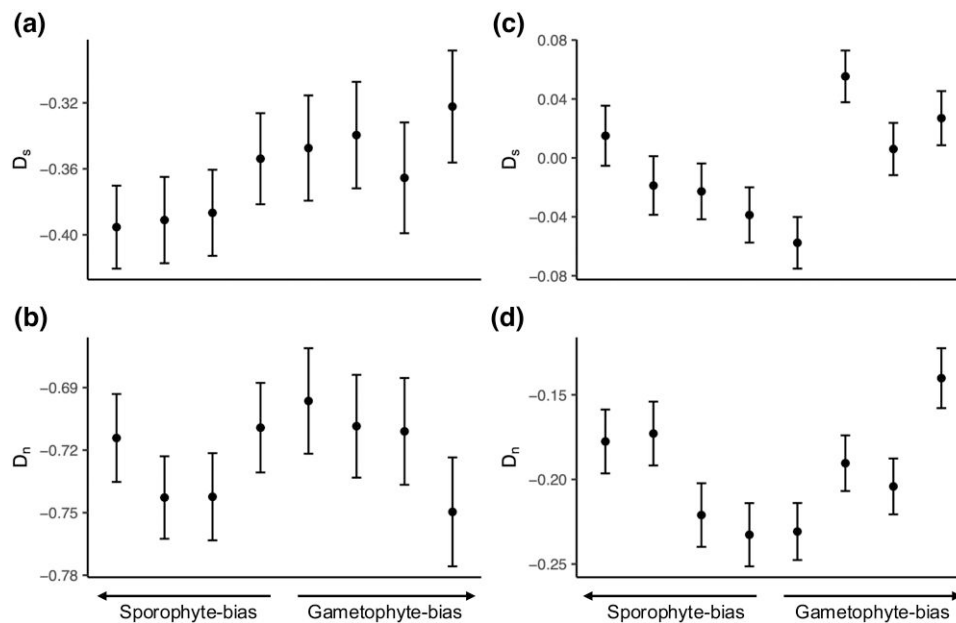


**Fig. 3.** Effect of expression level on weighted mean nucleotide diversity in *Rumex hastatulus* (a–c) and *Ceratodon purpureus* (d–f). Number of genes in each bin: 686 to 1,464 for *R. hastatulus*; 1,034 to 2,328 for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 500 bootstraps of the original dataset. qt1: lowest expression, qt10: highest expression, g: gametophyte-biased genes, s: sporophyte-biased genes. Note that the scales on the y axis are different between species.

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 (Fig. S6b). 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 (Fig. S7b). 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 (Fig. S6a), consistent with Fig. 2c. Across expression level quantiles, the confidence intervals mostly overlapped between gene sets and mutation categories in *R. hastatulus* (Fig. S7a).

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 ( $1 < N_{eS} < 10$ ) or strongly ( $N_{eS} > 10$ ) deleterious mutations, or a lower proportion of effectively neutral mutations ( $0 < N_{eS} < 1$ ). In contrast to previous studies (Arunkumar et al. 2013; Gossmann et al. 2014; Harrison et al. 2019; Gutiérrez-Valencia et al. 2022; 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 (Fig. S6). In the low to intermediate



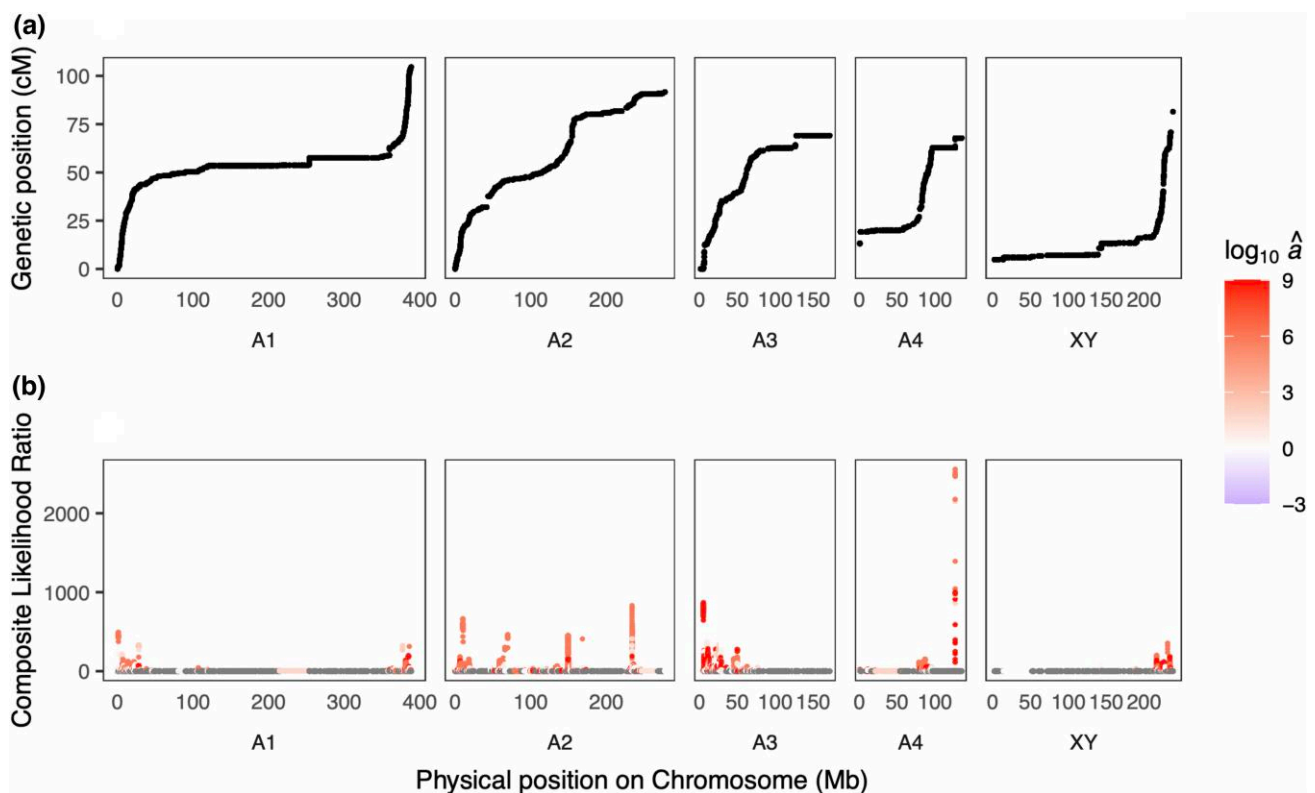
**Fig. 4.** Effect of expression bias between life stages on Tajima's  $D$  in *Rumex hastatulus* (a, b) and *Ceratodon purpureus* (c, d).  $x$  axis: quantiles of  $\log_2$ FoldChange between gametophyte and sporophyte expression. Number of genes in each bin: 1,009 to 1,734 ( $D_s$ ), 1,197 to 1,966 ( $D_r$ ) for *R. hastatulus*; 2,138 to 2,534 ( $D_s$ ), 2,097 to 2,579 ( $D_r$ ) for *C. purpureus*. Error bars represent mean  $\pm$  SEM across genes in each bin. Note that the scales on the  $y$  axis are different between species.

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 (Fig. S7). 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 to 70.21% in  $n$ , 65.38 to 68.66% in  $2n$ ; *C. purpureus*: 77.09 to 80.36% in  $n$ , 74.22 to 78.78% in  $2n$ , Fig. S7, Table S10).

### 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 (Fig. 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 (Fig. 5, Fig. S11). 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 (Fig. S8). 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  $\pi_r/\pi_s$  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  $\pi_r/\pi_s$  between genes under balancing



**Fig. 5.** Genome-wide scan of balancing selection in *Rumex 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.

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 (Figs. S9 and S10) 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 <  $2.2 \times 10^{-16}$ ). 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* are more prominent with a higher proportion of common alleles than in *R. hastatulus* (Figs. S9 and S10).

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 (Table S11). 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 (Tables S12 and S13). 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 (Table S12). 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 ( $\log_2\text{FoldChange} > 3$ , GO:0019953). These genes were also in the top 10% quantile of  $\pi_s$ .

## 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 (Fig. 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 (Fig. 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 (Fig. 3, Fig. S7). Gametophyte- and sporophyte-biased genes likely have more specialized functions and reduced

expression breadth than unbiased genes, particularly pollen-biased genes (Hony 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 (Table S5). 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 (Fig. 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 Materials 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 with 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. 2025). 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 (Tables S12 and S13). 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 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 et al. 2023), using standardized methods would help clarify how life stage complexity and duration affect the extent of overlapping and differential gene expression.

## Methods

### Plant Materials

For *R. hastatulus*, we collected 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. We collected pollen grains from the male  $F_1$  plants for RNA isolation based on a modified protocol from Lu (2011). We collected male leaf tissues for RNA isolation and female leaf tissues for DNA isolation using  $F_1$  plants, see [supplementary methods](#) for details. 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 to 15x. The median read numbers for DNA and RNA sequences are 72,907,022 (range: 64,585,743 to 88,888,320) and 23,412,838 (range: 18,618,763 to 58,151,564), respectively (Tables S1 and S2). 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 (Table S3). 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 to 64,278,575; Table S3).

### 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: (i) it was in the 10% or 90% quantiles of expression ratio among all significantly differentially expressed genes (adjusted  $P$ -value  $< 0.1$ ) and (ii) it had an absolute value of  $\log_2$ FoldChange  $> 2$ . 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 2022). We divided the genes into four bins based on quantiles of  $\log_2$ FoldChange for gametophyte-biased ( $\log_2$ FoldChange  $> 0$ ) and sporophyte-biased ( $\log_2$ FoldChange  $> 0$ ) genes separately. We defined unbiased genes as the middle two bins with lowest absolute value of  $\log_2$ FoldChange as well as an adjusted  $P$ -value  $\geq 0.1$ . 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](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](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 1,000 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.

### Supplementary Material

Supplementary material is available at [Genome Biology and Evolution online](#).

### Acknowledgments

We thank Bill Cole and Thomas Gludovacz for help with plant maintenance, Andrew Clemens for assistance with *C. purpureus* sporophyte cultivation, and Tyler Kent for help with interpolating recombination rate for *R. hastatulus*. This work was supported by NSERC Discovery Grants awarded to J.R.S., S.I.W., and S.C.H.B., and NSF and NASA awards to S.F.M. The *C. purpureus* sequencing (proposal: 10.46936/10.25585/60007219) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, was supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. We thank Vincent Castric and two anonymous reviewers for comments on the manuscript.

### Data Availability

Raw reads of *R. hastatulus* DNA and RNA Seq are available in NCBI SRA database under BioProject PRJNA744278. Raw reads of *C. purpureus* sporophyte RNA Seq are available through JGI genome portal, see project IDs in [Table S3](#). Scripts used in this study are available at [https://github.com/imengyuan/lifestage\\_conflict](https://github.com/imengyuan/lifestage_conflict).

### Literature Cited

- Alexa A, Rahnenfuhrer J. topGO: Enrichment analysis for gene ontology [Internet]. 2022. [accessed 2022 May 20]. Available from: <https://bioconductor.org/packages/topGO/>
- Andrews S. FastQC: a quality control tool for high throughput sequence data [Internet]. 2010. [accessed 2017 Dec 21]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 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*. 2013;30:2475–2486. <https://doi.org/10.1093/molbev/mst149>.
- Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI. Evolutionary genomics of plant gametophytic selection. *Plant Commun*. 2020;1:100115. <https://doi.org/10.1016/j.xplc.2020.100115>.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.

- Bitarello BD, Brandt DYC, Meyer D, Andrés AM. Inferring balancing selection from genome-scale data. *Genome Biol Evol.* 2023;15:evad032. <https://doi.org/10.1093/gbe/evad032>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Borg M, Brownfield L, Twell D. Male gametophyte development: a molecular perspective. *J Exp Bot.* 2009;60:1465–1478. <https://doi.org/10.1093/jxb/ern355>.
- Bowman JL, et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell.* 2017;171:287–304.e15. <https://doi.org/10.1016/j.cell.2017.09.030>.
- Carey SB, et al. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Sci Adv.* 2021a;7:eabh2488. <https://doi.org/10.1126/sciadv.abh2488>.
- Carey SB, et al. 2021c Mar 31. Novel insights into joint estimations of demography, mutation rate, and selection using UV sex chromosomes [Preprint]. *bioRxiv* 437085. <https://doi.org/10.1101/2021.03.30.437085>
- Carey SB, Kollar LM, Mcdaniel SF. Does degeneration or genetic conflict shape gene content on UV sex chromosomes? *Bryophyte Divers Evol.* 2021b;43:133–149. <https://doi.org/10.11646/bde.43.1.11>.
- Cervantes S, et al. Strong purifying selection in haploid tissue-specific genes of Scots pine supports the masking theory. *Mol Biol Evol.* 2023;40:msad183. <https://doi.org/10.1093/molbev/msad183>.
- Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta.* 2023;2:e107. <https://doi.org/10.1002/imt.2.107>.
- Cheng C, Kirkpatrick M. Sex-specific selection and sex-biased gene expression in humans and flies. *PLoS Genet.* 2016;12:e1006170. <https://doi.org/10.1371/journal.pgen.1006170>.
- Cheng X, DeGiorgio M. Detection of shared balancing selection in the absence of trans-species polymorphism. *Mol Biol Evol.* 2019;36:177–199. <https://doi.org/10.1093/molbev/msy202>.
- Cheng X, DeGiorgio M. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *Mol Biol Evol.* 2020;37:3267–3291. <https://doi.org/10.1093/molbev/msaa134>.
- Cheng X, DeGiorgio M. BalLeRMix+: mixture model approaches for robust joint identification of both positive selection and long-term balancing selection. *Bioinformatics.* 2022;38:861–863. <https://doi.org/10.1093/bioinformatics/btab720>.
- Clark AG. Sperm competition and the maintenance of polymorphism. *Heredity (Edinb).* 2002;88:148–153. <https://doi.org/10.1038/sj.hdy.6800019>.
- Crow JF, Kimura M. Evolution in sexual and asexual populations. *Am Nat.* 1965;99:439–450. <https://doi.org/10.1086/282389>.
- Danecek P, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Danecek P, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Dapper AL, Wade MJ. The evolution of sperm competition genes: the effect of mating system on levels of genetic variation within and between species. *Evolution.* 2016;70:502–511. <https://doi.org/10.1111/evo.12848>.
- Dobin A, et al. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics.* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Elyashiv E, et al. A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genet.* 2016;12:e1006130. <https://doi.org/10.1371/journal.pgen.1006130>.
- Field DL, Pickup M, Barrett SCH. The influence of pollination intensity on fertilization success, progeny sex ratio, and fitness in a wind-pollinated, dioecious plant. *Int J Plant Sci.* 2012;173:184–191. <https://doi.org/10.1086/663164>.
- Fishman L, Saunders A. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science.* 2008;322:1559–1562. <https://doi.org/10.1126/science.1161406>.
- Flintham E, Savolainen V, Otto SP, Reuter M, Mullon C. The maintenance of genetic polymorphism underlying sexually antagonistic traits. *Evol Lett.* 2025;9:259–272. <https://doi.org/10.1093/evlett/qrae059>.
- Friedman J, Barrett SCH. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Ann Bot.* 2009;103:1515–1527.
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Preprint]. *arXiv* 2012.1207.3907. <https://doi.org/10.48550/arXiv.1207.3907>
- Gaut BS, et al. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet.* 2007;8:77–84. <https://doi.org/10.1038/nrg1970>.
- Gerstein AC, Otto SP. Ploidy and the causes of genomic evolution. *J Hered.* 2009;100:571–581. <https://doi.org/10.1093/jhered/esp057>.
- Giner-Delgado C, et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. *Nat Commun.* 2019;10:4222. <https://doi.org/10.1038/s41467-019-12173-x>.
- Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol Biol Evol.* 2014;31:574–583. <https://doi.org/10.1093/molbev/mst226>.
- Gutiérrez-Valencia J, et al. Genomic signatures of sexual selection on pollen-expressed genes in *Arabidopsis alpina*. *Mol Biol Evol.* 2022;39:msab349. <https://doi.org/10.1093/molbev/msab349>.
- Haig D, Wilczek A. Sexual conflict and the alternation of haploid and diploid generations. *Philos Trans R Soc Lond B Biol Sci.* 2006;361:335–343. <https://doi.org/10.1098/rstb.2005.1794>.
- Haldane JBS. The causes of evolution. Princeton University Press; 1932.
- Harrison MC, Mallon EB, Twell D, Hammond RL. Deleterious mutation accumulation in *Arabidopsis thaliana* pollen genes: a role for a recent relaxation of selection. *Genome Biol Evol.* 2019;11:1939–1951. <https://doi.org/10.1093/gbe/evz127>.
- Healey AL, et al. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. *Nat Plants.* 2023;9:238–254. <https://doi.org/10.1038/s41477-022-01333-5>.
- Honys D, Twell D. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* 2003;132:640–652. <https://doi.org/10.1104/pp.103.020925>.
- Hormaza JJ, Herrero M. Pollen selection. *Theor Appl Genet.* 1992;83:663–672. <https://doi.org/10.1007/BF00226682>.
- Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. *Genetics.* 1987;116:153–159. <https://doi.org/10.1093/genetics/116.1.153>.
- Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. Exposing synonymous mutations. *Trends Genet.* 2014;30:308–321. <https://doi.org/10.1016/j.tig.2014.04.006>.
- Immler S. Haploid selection in “diploid” organisms. *Annu Rev Ecol Evol Syst.* 2019;50:219–236. <https://doi.org/10.1146/annurev-ecolsys-110218-024709>.
- Immler S, Arnqvist G, Otto SP. Ploidally antagonistic selection maintains stable genetic polymorphism. *Evolution.* 2012;66:55–65. <https://doi.org/10.1111/j.1558-5646.2011.01399.x>.
- Immler S, Otto SP. The evolutionary consequences of selection at the haploid gametic stage. *Am Nat.* 2018;192:241–249.
- Johnson MG, Shaw AJ. The effects of quantitative fecundity in the haploid stage on reproductive success and diploid fitness in the aquatic

- peat moss *Sphagnum macrophyllum*. *Heredity* (Edinb). 2016;116:523–530. <https://doi.org/10.1038/hdy.2016.13>.
- Kasimatis KR, Nelson TC, Phillips PC. Genomic signatures of sexual conflict. *J Hered*. 2017;108:780–790. <https://doi.org/10.1093/jhered/esx080>.
- Kidwell JF, Clegg MT, Stewart FM, Prout T. Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics*. 1977;85:171–183. <https://doi.org/10.1093/genetics/85.1.171>.
- King JL, Jukes TH. Non-darwinian evolution: most evolutionary change in proteins may be due to neutral mutations and genetic drift. *Science*. 1969;164:788–798.
- Klodová B, Fila J. A decade of pollen phosphoproteomics. *Int J Mol Sci*. 2021;22:12212. <https://doi.org/10.3390/ijms22212212>.
- Kondrashov AS, Crow JF. Haploidy or diploidy: which is better? *Nature*. 1991;351:314–315. <https://doi.org/10.1038/351314a0>.
- Korunes KL, Samuk K. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour*. 2021;21:1359–1368. <https://doi.org/10.1111/1755-0998.13326>.
- Kruglyak L, et al. Insufficient evidence for non-neutrality of synonymous mutations. *Nature*. 2023;616:E8–E9. <https://doi.org/10.1038/s41586-023-05865-4>.
- Lang D, et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J*. 2018;93:515–533. <https://doi.org/10.1111/tpj.13801>.
- Lankinen A, Kiboi S. Pollen donor identity affects timing of stigma receptivity in *Collinsia heterophylla* (Plantaginaceae): a sexual conflict during pollen competition? *Am Nat*. 2007;170:854–863. <https://doi.org/10.1086/522839>.
- Lawrie DS, Messer PW, Hershberg R, Petrov DA. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet*. 2013;9:e1003527. <https://doi.org/10.1371/journal.pgen.1003527>.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Lu Y. RNA isolation from *Arabidopsis* pollen grains. *Bio Protoc*. 2011;1:e67. <https://doi.org/10.21769/BioProtoc.67>.
- Mable BK, Otto SP. The evolution of life cycles with haploid and diploid phases. *BioEssays*. 1998;20:453–462.
- Mank JE. Population genetics of sexual conflict in the genomic era. *Nat Rev Genet*. 2017;18:721–730. <https://doi.org/10.1038/nrg.2017.83>.
- McDaniel SF. Divergent outcomes of genetic conflict on the UV sex chromosomes of *Marchantia polymorpha* and *Ceratodon purpureus*. *Curr Opin Genet Dev*. 2023;83:102129. <https://doi.org/10.1016/j.gde.2023.102129>.
- Ortiz-Ramírez C, et al. A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Mol Plant*. 2016;9:205–220. <https://doi.org/10.1016/j.molp.2015.12.002>.
- Otto SP, Scott MF, Immler S. Evolution of haploid selection in predominantly diploid organisms. *Proc Natl Acad Sci*. 2015;112:15952–15957. <https://doi.org/10.1073/pnas.1512004112>.
- Peters MAE, Weis AE. Selection for pollen competitive ability in mixed-mating systems. *Evolution*. 2018;72:2513–2536. <https://doi.org/10.1111/evo.13597>.
- Pickup M, Barrett SCH. The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. *Ecol Evol*. 2013;3:629–639. <https://doi.org/10.1002/ece3.465>.
- Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011;12:32–42. <https://doi.org/10.1038/nrg2899>.
- Qiu YL, Taylor AB, McManus HA. Evolution of the life cycle in land plants. *J Syst Evol*. 2012;50:171–194. <https://doi.org/10.1111/j.1759-6831.2012.00188.x>.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- Rifkin JL, et al. Widespread recombination suppression facilitates plant sex chromosome evolution. *Mol Biol Evol*. 2021;38:1018–1030. <https://doi.org/10.1093/molbev/msaa271>.
- Rifkin JL, et al. Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. *Philos Trans R Soc Lond B Biol Sci*. 2022;377:20210226. <https://doi.org/10.1098/rstb.2021.0226>.
- Rutley N, Twell D. A decade of pollen transcriptomics. *Plant Reprod*. 2015;28:73–89. <https://doi.org/10.1007/s00497-015-0261-7>.
- Ruzicka F, et al. The search for sexually antagonistic genes: practical insights from studies of local adaptation and statistical genomics. *Evol Lett*. 2020;4:398–415. <https://doi.org/10.1002/evl3.192>.
- Ruzicka F, et al. 2025 Feb 14. A century of theories of balancing selection [Preprint]. <https://doi.org/10.1101/2025.02.12.637871>.
- Sacchi B, et al. Phased assembly of neo-sex chromosomes reveals extensive y degeneration and rapid genome evolution in *Rumex hastatulus*. *Mol Biol Evol*. 2024;41:msae074. <https://doi.org/10.1093/molbev/msae074>.
- Sayadi A, et al. The genomic footprint of sexual conflict. *Nat Ecol Evol*. 2019;3:1725–1730. <https://doi.org/10.1038/s41559-019-1041-9>.
- Shen T, Gadiant P, Goodrich J, Becher H. 2024 Sep 9. Selection efficacy differs between lifestages in the haploid-diploid *Marchantia polymorpha* subsp. *ruderalis* [Preprint]. <https://doi.org/10.1101/2024.09.06.611587>.
- Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature*. 2022;606:725–731. <https://doi.org/10.1038/s41586-022-04823-w>.
- Shortlidge EE, et al. Microarthropod contributions to fitness variation in the common moss *Ceratodon purpureus*. *Proc Biol Sci*. 2021;288:20210119. <https://doi.org/10.1098/rspb.2021.0119>.
- Sigel EM, Schuettelpelz E, Pryer KM, Der JP. Overlapping patterns of gene expression between gametophyte and sporophyte phases in the fern *Polypodium amorphum* (Polypodiales). *Front Plant Sci*. 2018;9:1450. <https://doi.org/10.3389/fpls.2018.01450>.
- Slotte T, et al. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Genome Biol Evol*. 2011;3:1210–1219. <https://doi.org/10.1093/gbe/evr094>.
- Smith BW. The evolving karyotype of *Rumex hastatulus*. *Evolution*. 1964;93–104.
- Szövényi P, et al. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. *Mol Biol Evol*. 2013;30:1929–1939. <https://doi.org/10.1093/molbev/mst095>.
- Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ. Generation-biased gene expression in a bryophyte model system. *Mol Biol Evol*. 2011;28:803–812. <https://doi.org/10.1093/molbev/msq254>.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–595. <https://doi.org/10.1093/genetics/123.3.585>.

- Travers SE, Mazer SJ. Trade-offs between male and female reproduction associated with allozyme variation in phosphoglucosomerase in an annual plant (*Clarkia unguiculata*: Onagraceae). *Evolution*. 2001;55:2421–2428. <https://doi.org/10.1111/j.0014-3820.2001.tb00757.x>.
- Urrutia AO, Hurst LD. The signature of selection mediated by expression on human genes. *Genome Res*. 2003;13:2260–2264. <https://doi.org/10.1101/gr.641103>.
- Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE international parallel and distributed processing symposium (IPDPS). IEEE; 2019. p. 314–324.
- Walbot V, Evans MMS. Unique features of the plant life cycle and their consequences. *Nat Rev Genet*. 2003;4:369–379. <https://doi.org/10.1038/nrg1064>.
- Wright AE, et al. Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evol Lett*. 2018;2:52–61. <https://doi.org/10.1002/evl3.39>.
- Zhang J, Qian W. Functional synonymous mutations and their evolutionary consequences. *Nat Rev Genet*. 2025:1–16. <https://doi.org/10.1038/s41576-025-00850-1>.
- Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. *Nat Rev Genet*. 2015;16:409–420. <https://doi.org/10.1038/nrg3950>.
- Zhou Z, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci U S A*. 2016;113:E6117–E6125. <https://doi.org/10.1073/pnas.1606724113>.

Associate editor: Vincent Castric