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

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

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, CA
[2]Department of Biology, University of Florida, Gainesville, FL, USA
[3]Office of Research Development, University of Michigan, Ann Arbor, MI, USA
[4]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
[5]Department of Biology, Queen's University, Kingston, ON, Canada
[6]HudsonAlpha Genome Sequencing Center, Huntsville, AL, USA
*These authors made equal contributions.
†Corresponding author: E-mail: my.yuan@mail.utoronto.ca

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 (2*n*) and gametophytes (*n*) may undergo differential or even opposing selection. Theoretical work suggests the 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 among life stages and between 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.

Keywords: balancing selection, antagonistic pleiotropy, intralocus conflict, gene expression, gametophytic selection

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, hereafter life stages (Kidwell et al. 1977; Immler et al. 2012). Despite persistent theoretical and empirical interests, 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 focus on testing 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 and 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

73 selection on gametophytes (gametophytic selection) affects sporophytes is a key unresolved
74 question in plant evolutionary genetics.
75
76 The evolutionary consequences of gametophytic selection have been examined both theoretically
77 and empirically (Haldane 1932; Otto et al. 2015; Immler and Otto 2018; summarized in Immler
78 2019, Beaudry et al. 2020). The haploid nature of the gametophytic phase is predicted to increase
79 the efficacy of selection due to the lack of dominance (i.e., the masking hypothesis), allowing for
80 more efficient removal of recessive deleterious mutations (i.e., purging) and fixation of recessive
81 beneficial mutations (Crow and Kimura 1965; Kondrashov and Crow 1991; Gerstein and Otto
82 2009). Consistent with this prediction, some studies have found stronger purifying and positive
83 selection in gametophyte-specific than sporophyte-specific genes (Arunkumar et al. 2013;
84 Cervantes et al. 2023; but see Harrison et al. 2019). Stronger selection on gametophyte-expressed
85 genes can also be driven by gametophytic competition (sexual selection) in angiosperms
86 (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et al. 2022). Additionally,
87 gene expression breadth can be a stronger predictor of selection efficacy than the masking effect
88 in angiosperms and bryophytes (Szövényi et al. 2013).
89
90 The common genome shared between life stages provides an opportunity to test how
91 gametophytic selection affects the evolution of the sporophyte stage due to pleiotropy between
92 life stages. Plants share a substantial overlap in gene expression between life stages based on both
93 microarray and mRNA sequencing studies (summarized in Beaudry et al. 2020). In ferns and
94 bryophytes where the gametophytes are photosynthetic and exhibit indeterminate growth, gene
95 expression overlap may exceed that in angiosperms, e.g., estimates are ~60% in the angiosperm
96 *Arabidopsis thaliana* (Walbot and Evans 2003; Borg et al. 2009), 97.7% in the fern *Polypodium*

97 *amorphum* (Sigel et al. 2018), and 85% in the bryophyte *Physcomitrium patens* (Ortiz-Ramírez et
98 al. 2016). Previous studies on gametophytic selection have compared selective forces in
99 gametophyte -*vs.*- sporophyte-specific genes to test for the effects of haploid expression on the
100 distribution of selection coefficients (Arunkumar et al. 2013; Gossmann et al. 2014; Harrison et
101 al. 2019; Gutiérrez-Valencia et al. 2021; Cervantes et al. 2023). However, the patterns of genetic
102 diversity and selection pressures of genes with overlapping expression, which are more likely to
103 be pleiotropic due to the greater potential for competing selection pressures across life stages
104 (Immler et al. 2012; Peters and Weis 2018), have not yet been specifically examined.
105
106 In theory, antagonistic pleiotropy between life stages, i.e., intralocus conflict when alleles of a
107 gene have opposite fitness effects between life stages, can cause balancing selection and maintain
108 genetic variation under some conditions (Immler et al. 2012; Peters and Weis 2018). Despite the
109 theoretical support, empirical evidence of antagonistic pleiotropy in angiosperms is limited to a
110 few studies suggesting sexual conflict across life stages, e.g., between male gametophyte and
111 female sporophyte during pollen competition (e.g., Travers and Mazer 2001; Lankinen and Kiboi
112 2007), and between female gametophyte and male sporophyte in a female meiotic drive system
113 (Fishman and Saunders 2008). In bryophytes where sporophytes are dependent only on female
114 gametophytes, parent offspring conflict may occur (Haig and Wilczek 2006; Johnson and Shaw
115 2016; Shortlidge et al. 2021), but the mechanism and genetic architecture underlying any such
116 conflict remain unclear. Testing the genome-wide prevalence of antagonistic pleiotropy between
117 life stages and its potential to maintain genetic variation is therefore key to understanding the
118 genetic architecture of conflict between plant life stages.
119

120 Given the potential for antagonistic pleiotropy between the life stages for genes with overlapping
121 expression, we predicted that if these effects are widespread, we should see population genetic
122 signals of balancing selection more commonly in genes with overlapping expression (Immler et
123 al. 2012; Peters and Weis 2018). An analogous finding is that genes with intermediate rather than
124 extreme sex bias between sexes show patterns of genetic diversity consistent with balancing
125 selection, as expected if extreme expression bias reflects resolved rather than ongoing conflict
126 (e.g., Cheng and Kirkpatrick 2016; Wright et al. 2018; Sayadi et al. 2019; summarized in Ruzika
127 et al. 2020). Previous studies on sexual conflict and sex-biased expression studies primarily relied
128 on diversity statistics. Recently developed model-based approaches offer an additional way to
129 detect balancing selection (Bitarello et al. 2023) and assess enrichment of biased genes between
130 life stages.
131
132 We examined gene expression and genetic diversity patterns in the gametophytic and sporophytic
133 life stages of two plant species: the angiosperm *Rumex hastatulus* (Polygonaceae) and the
134 bryophyte *Ceratodon purpureus* (Ditrichaceae). The annual *R. hastatulus* is dioecious, obligately
135 outcrossing, and wind-pollinated with an XY sex chromosome system (Smith 1964; Rifkin et al.
136 2021). The uniovulate flowers and the large quantity of pollen commonly produced in wind-
137 pollinated angiosperms suggests high potential for gametophytic selection during pollen
138 competition (Friedman and Barrett 2009; Field et al. 2012; Immler 2019). In the dioicous moss
139 *C. purpureus*, the female and male gametophytes possess the U or V sex chromosome,
140 respectively (Carey et al. 2021a), they produce gametes that fuse to form a diploid zygote (i.e.,
141 sporophyte, unsexed); the sporophyte matures and ultimately undergoes meiosis, but its entire
142 development occurs on the female gametophyte plant, providing multiple opportunities for
143 antagonistic pleiotropy (Carey et al. 2021b; McDaniel 2023). The contrasts in the length and

144 complexity of the gametophytic stage in these two species allow us to evaluate the evolutionary
145 consequences of biphasic plant life cycles in a broader context.
146
147 We first compared gene expression in gametophytic and sporophytic tissues of *R. hastatulus* and
148 *C. purpureus,* finding a larger overlap in gene expression and a smaller proportion of biased
149 genes between life stages in *C. purpureus* than *R. hastatulus*. We then examined the population
150 genetic signals of balancing selection due to antagonistic pleiotropy between life stages by
151 examining nucleotide diversity and Tajima's *D* across varying degrees of expression bias
152 between life stages. In contrast with our predictions for genome-wide patterns of antagonistic
153 pleiotropy, our results show elevated nucleotide diversity and Tajima's *D* in highly gametophyte-
154 biased genes in both *R. hastatulus* and *C. purpureus*. We conclude that expression levels and
155 breadth were a stronger predictor of selection efficacy than potential competing selective
156 pressures between life stages. Lastly, we identified in both species genomic regions containing
157 hundreds of candidate genes under balancing selection using a composite likelihood ratio test
158 (Cheng and DeGiorgio 2019, 2020, 2022) and examined the biological functions and differential
159 gene expression patterns of these candidate genes.
160
161 **Results**
162 *Gene expression in gametophytes and sporophytes*
163 We compared gene expression in gametophyte and sporophyte tissues of *R. hastatulus* and *C.*
164 *purpureus*. In *R. hastatulus*, we sequenced 75 mature pollen samples and 77 male leaf samples
165 representing the gametophyte and sporophyte stages, respectively (Table S1). We used only male
166 leaf tissue for assessing sporophytic expression in *R. hastatulus* to avoid confounding effects
167 from potential sexual conflict between leaves of males and females. In *C. purpureus*, we

168 compared expression between 21 sporophyte and 36 gametophyte samples (whole plant,
169 described in Carey et al. 2021a, Table S3, see Materials and Methods). In both species, there is a
170 large overlap in gene expression between the life stages (Table 1). In *R. hastatulus*, we identified
171 15,958 gametophyte-expressed genes, consisting of 52% of all annotated genes across the
172 genome, and 89% of genes that were expressed in either life stage. In *C. purpureus*, we identified
173 24,097 gametophyte-expressed genes, consisting of 69% of all annotated genes and 93% of genes
174 that were expressed in either life stage. The overlap in gene expression between life stages (89%
175 in *R. hastatulus*, and 93% in *C. purpureus*) sets up a strong potential for pleiotropic effects of
176 mutations between life stages.
177
178 We first identified genes that were differentially expressed (DE) between life stages and
179 exclusively expressed in one life stage (Table S4). There was a slightly higher percentage of
180 overlapping gene expression and a lower percentage of differentially expressed genes between
181 life stages in *C. purpureus* than *R. hastatulus* (Table 1). In *R. hastatulus*, we identified 5,635
182 gametophyte-biased and 8,653 sporophyte-biased genes; similarly in *C. purpureus*, there were
183 6,446 gametophyte-biased and 5,202 sporophyte-biased genes (Table 1). Among significant DE
184 genes, we filtered for genes in the top 10% quantile of expression bias, allowing us to compare
185 genes with different tissue specificities. Our filtering process resulted in similar numbers of
186 gametophyte- and sporophyte-specific genes: 1,711 gametophyte-specific and 1,711 sporophyte-
187 specific genes in *R. hastatulus*, and 1,886 gametophyte-specific and 1,867 sporophyte-specific
188 genes in *C. purpureus*. Gene ontology (GO) enrichment analyses indicated very different
189 functional enrichment for biased and specific genes between life stages (Table S5, S6),
190 highlighting distinctive expression profiles across life stages in both species. For example, in *R.*
191 *hastatulus*, gametophyte-specific genes had protein phosphorylation functions (GO:0006468)

192 which are important during pollen-tube growth (Klodová and Fíla 2021) and sporophyte-specific

193 genes showed photosynthetic functions (GO:0015979).

194

195 Table 1. Summary of samples used and gene expression analyses in *R. hastatulus* and *C.*

196 *purpureus*. *P*-adj: Benjamini-Hochberg adjusted *p*-value with a false discovery rate of 0.1, FC:

197 fold change in expression, i.e., the ratio of expression level between the two tissues compared.

198 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%)

199

200

201 *Nucleotide diversity and expression bias between life stages*

202 Population genomic analyses can be used to provide an indirect test for pleiotropic effects across

203 life stages. We calculated nucleotide diversity at 4-fold (π_s) and 0-fold (π_n) sites for each gene and

204 looked at the linear relation between π_s or π_n and the direction and degree of expression bias

205 between life stages (log2FoldChange) across genes. The linear regression of π_s or π_n on the

206 direction and degree of expression bias between life stages (log2FoldChange and its absolute

207 value, where positive values indicate gametophyte bias and negative values indicate sporophyte-

208 bias) showed a significant but weak positive relationship, with near-zero slopes in both species

209 (Figure S1, S2). We next divided the genes into eight bins based on the direction and degree of

210 expression bias between life stages and calculated the weighted mean of π_s and π_n for each bin,
211 weighted by the number of sites per gene. In both species, highly gametophyte-biased genes
212 exhibited the highest levels of π_s, π_n and π_n/π_s ratio (Figure 1, Table S7). This result contrasts with
213 our expectation that genes with overlapping and intermediate bias in expression should show the
214 highest diversity under a model of widespread antagonistic pleiotropy. In both species, π_s and π_n
215 increase with the degree of expression bias between life stages in both directions (less so in *R.*
216 *hastatulus*), but values of π_s and π_n were greatest for the highly gametophytic-biased genes
217 (Figure 1). In *R. hastatulus*, the increase in π_n with the degree of expression bias was less
218 symmetrical than in π_s, with sporophyte-biased genes having lower π_n than genes with
219 overlapping expression (Figure 1a, 1b). Our results were robust to our choice of weighting π_s and
220 π_n by the number of sites per gene, as when we estimated the mean and standard errors without
221 weighting by the number of sites, the patterns remained the same (Figure S3).
222
223 There are two possible explanations for the patterns of elevated π_s and π_n in the most
224 gametophyte-biased genes, a greater prevalence of balancing selection under the assumption of
225 neutrality of synonymous mutations or relaxed purifying selection on this set of genes. If there is
226 relaxed purifying selection on gametophytic-biased genes, this could drive higher π_n and higher
227 π_n/π_s. The observed π_n/π_s ratios showed a similar trend as π_s and π_n, especially in *C. purpureus*,
228 with highly gametophyte-biased genes having the highest values (Figure 1c, 1f), suggesting either
229 balancing selection, or a reduction in the strength of purifying selection on the most
230 gametophyte-biased genes.

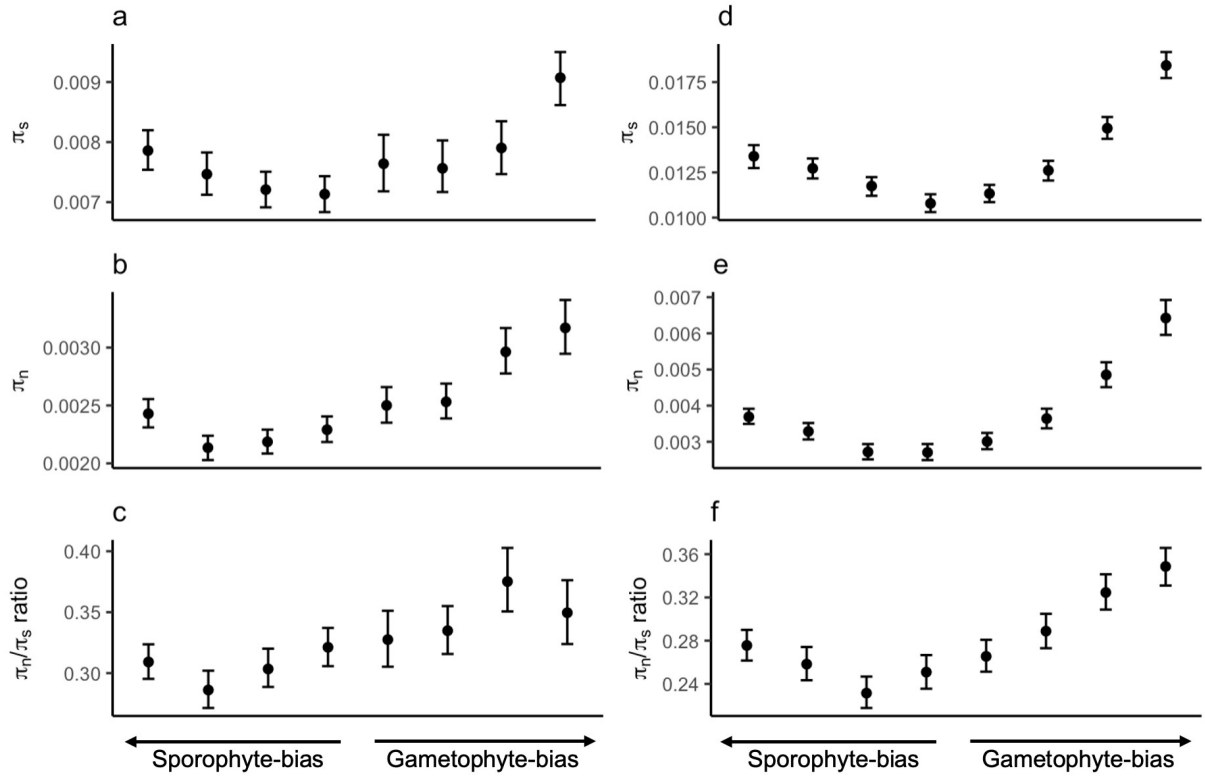
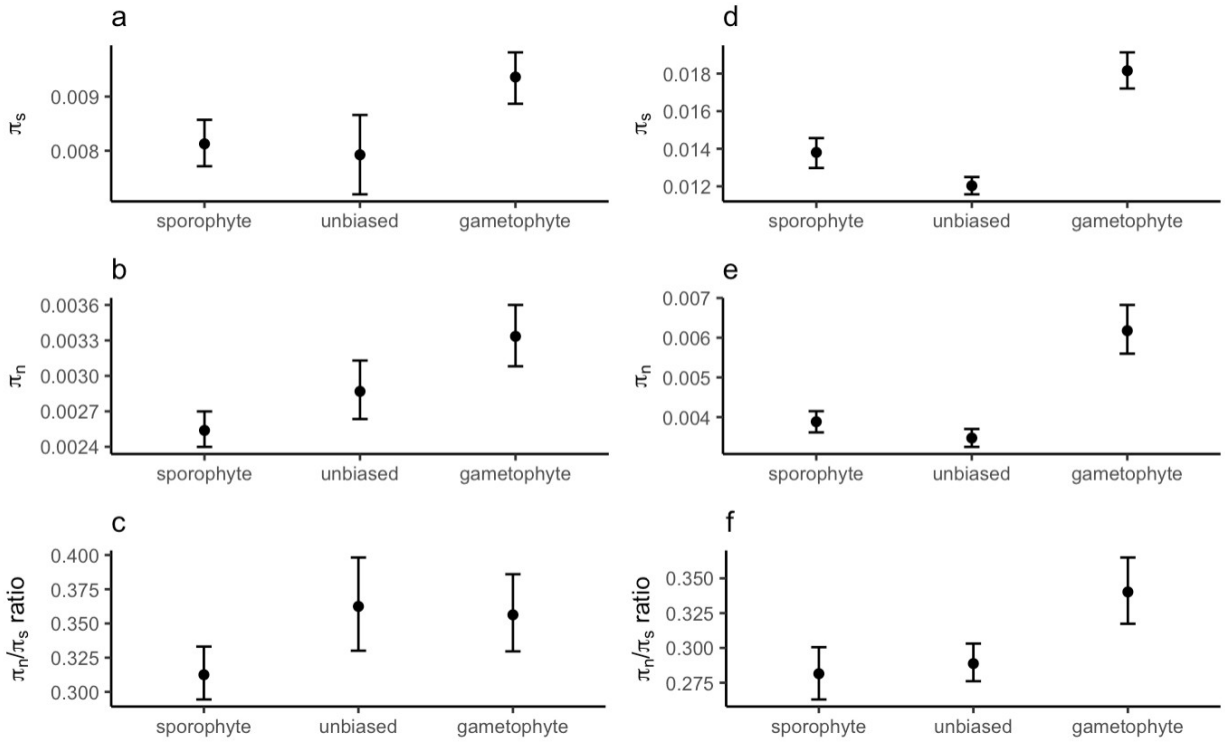


Figure 1. Effect of expression bias between life stages on nucleotide diversity in *R. hastatulus* (a-c) and *C. purpureus* (d-f). x-axis: quantiles of log2FoldChange between gametophyte and sporophyte expression. Number of genes in each bin: 1,009-1,847 for *R. hastatulus*; 1,943-2,564 for *C. purpureus*. 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 π_s, π_n and π_n/π_s ratio for gametophyte-specific, sporophyte-specific and unbiased genes (Figure 2, S4; Table S8). The results are mostly consistent with Figure 1, with gametophyte-specific genes having a higher level of π_s and π_n than

244 sporophyte-specific and unbiased genes in both species (Figure 2a, 2b, 2d, 2e), although the π_n/π_s

245 ratios were not elevated above unbiased genes in *R. hastatulus* as in *C. purpureus* (Figure 2c, 2f).



246

247 Figure 2. Weighted mean nucleotide diversity of gametophyte-specific, sporophyte-specific,

248 unbiased genes in *R. hastatulus* (a-c) and *C. purpureus* (d-f). Number of genes in each category:

249 1149 (s), 920 (g), 524 (unbiased) for *R. hastatulus*; 1,232 (s), 1,217 (g), 3380 (unbiased) for *C.*

250 *purpureus*. The weighted mean and 95% confidence intervals are based on 1000 bootstraps of the

251 original dataset. Note that the scales on the y-axis are different between species.

252

253 One possible additional factor that could contribute to differences in the strength of purifying

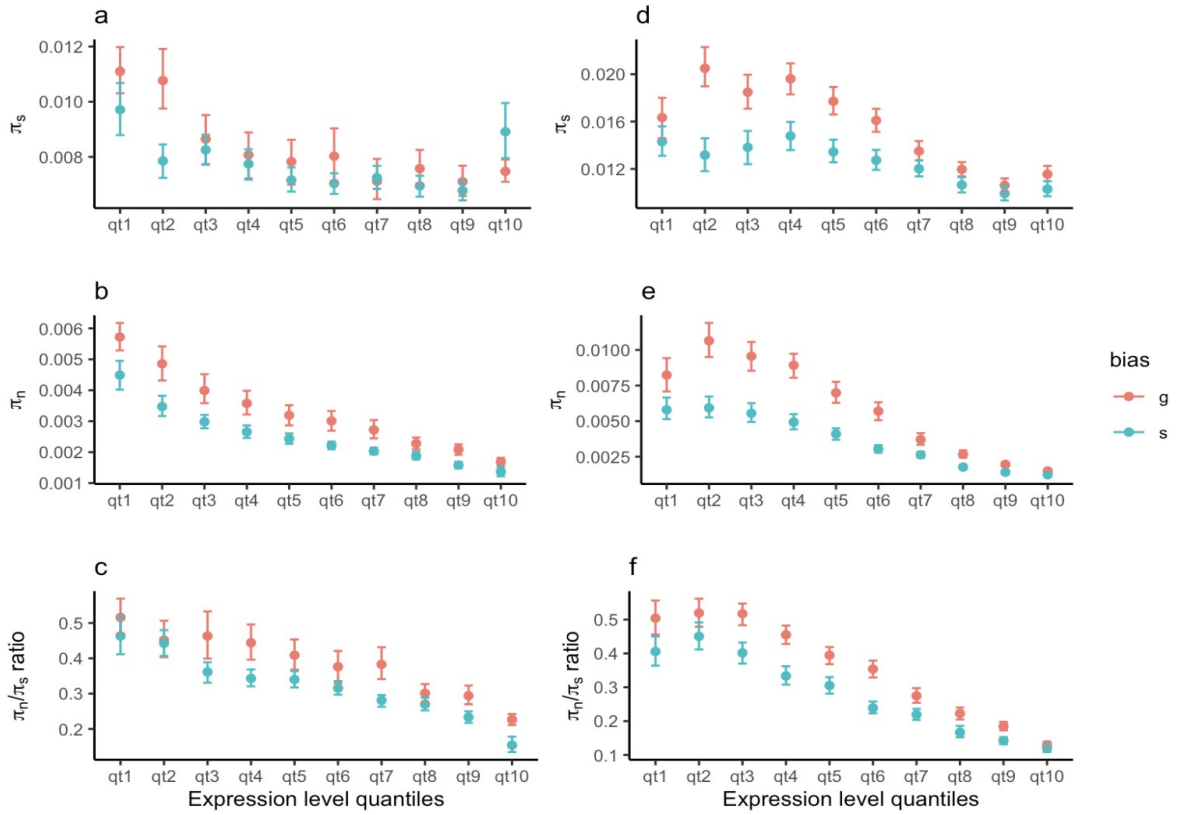
254 selection is if these genes differ in expression level. Gene expression has been shown to be a

255 strong predictor of the strength of purifying selection genome-wide in many species (e.g., Urrutia

256 and Hurst 2003; Slotte et al. 2011; Zhang and Yang 2015). We used this relation to further

257 examine the possible effects of relaxed purifying selection on π_s, and divided the genes into ten

258 bins of expression level (baseMean across all samples) to examine how expression level affects
259 π_s, π_n and π_n/π_s in gametophyte- and sporophyte-biased genes. As expression level increases, both
260 π_n and π_s decrease, with a bigger decrease in π_n than π_s (Figure 3). Gametophyte-biased genes
261 have higher π_s, π_n and π_n/π_s than sporophyte-biased genes, regardless of expression level in both
262 species, except for the bins with overlapping confidence intervals in π_s of *R. hastatulus* (Figure
263 3). Our results suggest that expression level has a strong effect on patterns of diversity. Overall,
264 contrary to our expectations, genes with overlapping expression levels did not show signals of
265 elevated diversity due to balancing selection. Instead, heterogeneity in diversity likely reflects
266 differences in the strength of purifying selection.



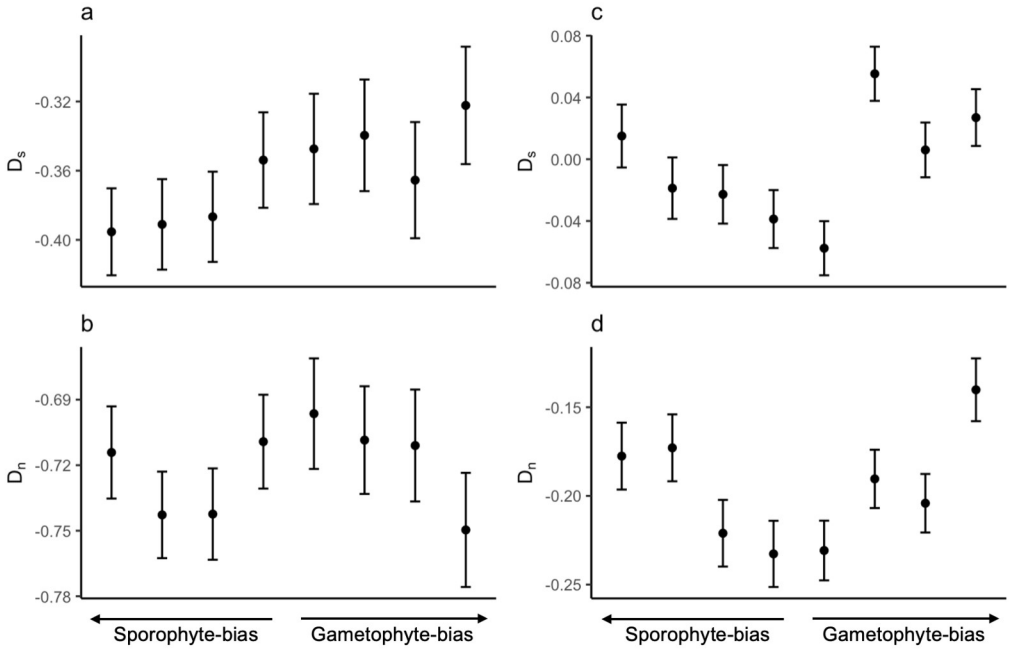
267

268 Figure 3. Effect of expression level on weighted mean nucleotide diversity in *R. hastatulus* (a-c)
269 and *C. purpureus* (d-f). Number of genes in each bin: 686-1,464 for *R. hastatulus*; 1,034-2,328
270 for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 500 bootstraps

271 of the original dataset. qt1: lowest expression, qt10: highest expression. Note that the scales on
272 the y-axis are different between species.
273
274 *Tajima's D and expression bias between life stages*
275 Tajima's *D* is a metric summarizing information of the distribution of allele frequencies in a
276 population (Tajima 1989) and is influenced by both selection and demographic history. Positive
277 values are associated with balancing selection or population contraction and negative values
278 indicate purifying selection or population expansion. We further tested the effect of expression
279 bias between life stages on Tajima's *D* at synonymous (D_s) and nonsynonymous (D_n) sites
280 (Figure 4). In both species, D_n was smaller than D_s, consistent with purifying selection on
281 nonsynonymous sites (Figure 4a v.s. 4b, 4c v.s. 4d). In *R. hastatulus*, both D_n and D_s were
282 negative, D_s was higher in gametophyte-biased genes than unbiased and sporophyte-biased genes,
283 as expected if they had a higher proportion of genes under balancing selection (Figure 4a). The
284 negative values in D_n and D_s in *R. hastatulus* may possibly be attributed to population expansion,
285 but we currently have no evidence on demographic history. In *C. purpureus*, the patterns of D_n
286 were more similar to π_s and π_n than D_s, both D_s and D_n were higher in highly gametophyte-biased
287 genes than unbiased genes (Figure 4c, 4d). The higher and more positive Tajima's *D* values in *C.*
288 *purpureus* compared to *R. hastatulus* likely resulted from differences in the two species'
289 demographic histories. We also compared D_s and D_n in gametophyte- and sporophyte-biased
290 genes across expression level quantiles. In *R. hastatulus*, D_s and D_n remained roughly constant
291 across expression level quantiles, whereas in *C. purpureus*, D_s and D_n decreased with higher
292 expression (Figure S5). Gametophyte- and sporophyte-biased genes had similar levels of D_s and
293 D_n across expression levels in both species. Combined with patterns of π (Figure 1-3) our results

294 are consistent with the hypothesis that gametophyte-biased genes are under relaxed purifying

295 selection.



296

297 Figure 4. Effect of expression bias between life stages on Tajima's *D* in *R. hastatulus* (a, b) and

298 *C. purpureus* (c, d). x-axis: quantiles of log2FoldChange between gametophyte and sporophyte

299 expression. Number of genes in each bin: 1,009-1,734 (D_s), 1,197-1,966 (D_n) for *R. hastatulus*;

300 2,138-2,534 (D_s), 2,097-2,579 (D_n) for *C. purpureus*. Error bars represent mean ± SEM across

301 genes in each bin. Note that the scales on the y-axis are different between species.

302

303 *Distribution of fitness effects*

304 Tajima's *D* values reflect a summary of the site frequency spectrum rather than an explicit

305 quantification of the strength of purifying selection on nonsynonymous sites. To address this, we

306 estimated the distribution of fitness effect (DFE) to further understand the strength of purifying

307 selection across life stages while controlling for expression level (Supplementary Methods). We

308 first compared the DFE of gametophyte-specific, sporophyte-specific and unbiased genes (Figure

309 S6, Table S9), then we binned the gene sets into four expression level quantiles (Figure S7, Table
310 S10). In both species, higher expression level was associated with lower proportions of
311 effectively neutral mutations ($0 < N_e s < 1$), consistent with an increasing level of constraint due to
312 higher expression level (Figure S7).
313
314 In *C. purpureus*, unbiased genes had a lower proportion of effectively neutral mutations and a
315 higher proportion of strongly deleterious mutations than gametophyte- or sporophyte-specific
316 genes. This finding is consistent with unbiased genes experiencing higher levels of constraint
317 whereas stage-specific genes are more likely under relaxed purifying selection (Figure S6b).
318 Similarly, in the highest expression level quantile (qt4), unbiased genes had a slightly lower
319 proportion of effectively neutral mutations than gametophyte- or sporophyte-specific genes.
320 However, in the other expression level quantiles the same pattern was not evident: unbiased
321 genes had a higher proportion of effectively neutral mutations in the first quantile, and lower
322 proportion of strongly deleterious mutations in the second quantile (Figure S7b). In *R. hastatulus*,
323 gametophyte-specific and unbiased genes had similar proportions of effectively neutral and
324 strongly deleterious mutations and sporophyte-specific genes had the lowest proportion of
325 effectively neutral mutations and the highest proportion of strongly deleterious mutations (Figure
326 S6a), consistent with Figure 2c. Across expression level quantiles, the confidence intervals
327 mostly overlapped between gene sets and mutation categories in *R. hastatulus* (Figure S7a).
328
329 Based on the haploid purging hypothesis, we predicted that selection efficacy would be stronger
330 in haploid than diploid phases. We therefore expected gametophyte-specific genes to have higher
331 proportions of slightly ($1 < N_e s < 10$) or strongly ($N_e s > 10$) deleterious mutations, or a lower
332 proportion of effectively neutral mutations ($0 < N_e s < 1$). In contrast to previous studies

333 (Arunkumar et al. 2013; Gossmann et al. 2014; Harrison et al. 2019; Gutiérrez-Valencia et al.
334 2021; Cervantes et al. 2023), our results in both species did not consistently support the haploid
335 purging hypothesis. In *R. hastatulus*, gametophyte-specific genes had a higher proportion of
336 strongly deleterious mutations than sporophyte-specific genes, while in *C. purpureus* the
337 proportions were similar with overlapping confidence intervals (Figure S6). In the low to
338 intermediate expression quantiles (qt1 to qt3), sporophyte-specific genes showed a slightly higher
339 proportion of strongly deleterious mutations ($N_e s > 10$) than gametophyte-specific genes in both
340 species (Figure S7). In the highest expression quantile (qt4), gametophyte-specific genes had
341 slightly higher proportions of strongly deleterious mutation ($N_e s > 10$) than sporophyte-specific
342 genes in both species, but the differences were not significant with overlapping confidence
343 intervals (*R. hastatulus*: 67.19 - 70.21% in n, 65.38 - 68.66% in 2n; *C. purpureus*: 77.09 - 80.36%
344 in n, 74.22 - 78.78 % in 2n, Figure S7, Table S10).
345
346 *Genome-wide scan for balancing selection*
347 These overall patterns of diversity do not show a genome-wide signal of an enrichment of
348 balancing selection on genes with overlapping expression between the life stages; however, it is
349 possible that individual genes may still be subject to balancing selection. We therefore took an
350 alternative approach by conducting a genome-wide scan for signals of balancing selection and
351 explored whether they were enriched for genes that might be subject to antagonistic pleiotropy
352 between life stages. We applied a composite likelihood ratio test to scan the genome for
353 candidate genes under balancing selection based on B_2 statistics that account for both increased
354 polymorphism density and an excess of intermediate-frequency alleles (Cheng and DeGiorgio
355 2019, 2020, 2022). In *R. hastatulus,* a total number of 4,945,561 sites were tested in windows
356 based on genetic positions (Figure 5). We found 79,431 sites with strong signals of balancing

357 selection across all chromosomes (CLR > 9.5), located at 449 genes, 405 of which had evidence
358 of expression. The candidate genes were only located in regions with considerable recombination
359 where genetic positions increased rapidly with physical positions, based on our genetic maps
360 (Figure 5, Figure S11). However, among the genes found in regions with evidence of balancing
361 selection, 138 were located within inversion polymorphisms identified from two phased whole
362 genome assemblies of *R. hastatulus* on chromosomes A2 (63), A3 (74), and A4 (1) (Sacchi et al.
363 2024; Sacchi et al, in prep), consisting of more than 30% of all candidate genes; whereas these
364 inversions contained 16.7% of genes across the genome. Because inversion heterozygotes are
365 likely to experience reduced recombination, these candidate genes may be subject to non-
366 independent signals of balanced polymorphism. In *C. purpureus*, the scans were performed in
367 physical position windows as site-specific recombination rates were unavailable (Figure S8). We
368 tested 7,743,902 sites and identified 339,375 sites under balancing selection (CLR > 956.2),
369 located at 963 genes, 807 of which had evidence of expression in this study. In *R. hastatulus*, as
370 expected candidate genes under balancing selection had significantly higher π_s, and significantly
371 lower π_n/π_s than genes that were not under balancing selection (Welch Two Sample *t*-test, *p*-value
372 = 0.0001511, 0.0003204, respectively). In *C. purpureus*, there was no significant differences in π_s
373 or π_n/π_s between genes under balancing selection and gene that were not (Welch Two Sample *t*-
374 test, *p*-value = 0.6029; Two Sample *t*-test, *p*-value = 0.5956, respectively). We compared the site
375 frequency spectrum (SFS) of all sites tested and the sites under balancing selection (Figure S9,
376 S10) and found sites under balancing selection had significantly higher minor allele frequencies
377 than the input data in both species (Mann-Whitney *U* test, *p*-value < 2.2e-16). Note that this is
378 expected, given the use of the site frequency spectrum in the model-based inference. The
379 differences in the shape of SFS in *C. purpureus* is more prominent with a higher proportion of
380 common alleles than in *R. hastatulus* (Figure S9, S10).

381

382 We performed contingency tests on the number of genes in different expression bias categories
383 (gametophyte-biased, sporophyte-biased, unbiased) for genes under balancing selection and
384 across the whole genome (Tables S11). In both species, we found no enrichment of any category
385 of expression bias under long-term balancing selection compared to the whole genome (Fisher's
386 exact test, $p > 0.05$). We examined the gene functions for all gametophyte-biased genes in
387 regions under balancing selection (Table S12, S13). We performed GO enrichment after
388 excluding the genes found within inversion polymorphisms in *R. hastatulus*, since the
389 suppression of recombination between inversion types can influence our ability to find functional
390 enrichments of causal loci. The GO enrichment results after excluding inversions were not
391 compelling (Table S12). In *R. hastatulus*, two genes under balancing selection had probable
392 disease resistance functions based on orthologs to *Arabidopsis thaliana* (At4g19060,
393 At4g14610). In *C. purpureus*, two genes under balancing selection were expressed in both life
394 stages with high gametophyte bias and have sexual reproduction functions (log2FoldChange > 3,
395 GO:0019953). These genes were also in the top 10% quantile of π_s.

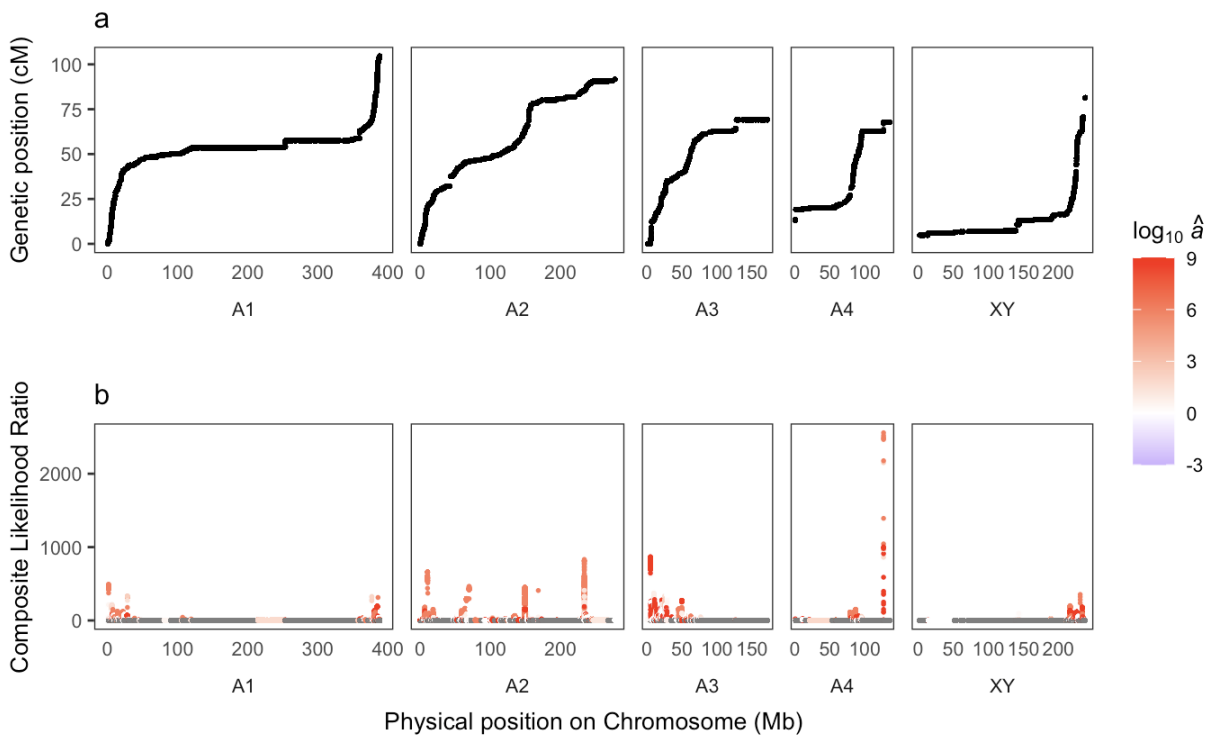


Figure 5. Genome-wide scan of balancing selection in *R. hastatulus*. y-axis: genetic positions of sites being tested (a), composite likelihood ratio (b). $\hat{a}$: estimated dispersion parameter, a positive $\log_{10}\hat{a}$ value suggests balancing selection.

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 π_s, π_n and D_n in highly gametophyte-biased genes in both species. Combined with the observations that π_s, π_n and π_n/π_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

410 hundreds of candidate genes under balancing selection, but these genes were not enriched for
411 gametophyte-biased expression. Below we discuss how these findings inform our understanding
412 of plant gametophytic selection and the challenges of alternative approaches for detecting
413 balancing selection.
414
415 *Widespread relaxed purifying selection*
416 Balancing selection driven by sexual antagonism has been tested using genetic diversity and sex-
417 biased expression (Cheng and Kirkpatrick 2016; Kasimatis et al. 2017; Mank 2017; Wright et al.
418 2018; Sayadi et al. 2019), showing elevated diversity in weakly sex-biased genes consistent with
419 ongoing conflict. We predicted similar patterns for life stages, however, we found higher π_s, π_n
420 and D_n in highly gametophyte-biased genes (Figure 1), which could result from balancing
421 selection or relaxed purifying selection. As highly expressed genes are under strong selective
422 constraint (e.g., Urrutia and Hurst 2003; Slotte et al. 2011; Zhang and Yang 2015), high π_s in
423 gametophyte-biased genes with high expression level would suggest balancing selection. Since π_s
424 decreased with higher expression level (Figure 3), the elevated π_s in gametophyte-biased genes
425 suggests relaxed constraint on both nonsynonymous and synonymous sites (Zhang and Qian
426 2025) is the more plausible explanation rather than widespread intralocus conflict.
427
428 Higher π_s under relaxed purifying selection could stem from weaker background selection and/or
429 less direct selection on synonymous sites themselves. Background selection reduces neutral
430 nucleotide diversity due to linkage to deleterious mutations, with greater reductions in more
431 constrained regions (Elyashiv et al. 2016). Highly gametophyte-biased genes may be under
432 weaker background selection, elevating π_s and π_n. Though synonymous mutations are generally
433 assumed to be neutral (King and Jukes 1969), studies in fruit flies and yeast showed many

434 synonymous mutations can be deleterious (Lawrie et al. 2013; Shen et al. 2022; Zhang and Qian
435 2025; but see Kruglyak et al. 2023), likely due to codon usage bias (Plotkin and Kudla 2011;
436 Hunt et al. 2014; Zhou et al. 2016). In *Marchantia polymorpha*, relaxed purifying selection and
437 lower codon usage bias were found in sporophyte-biased genes (Shen et al. 2024). In our case,
438 both synonymous and nonsynonymous sites may be constrained but less so in highly
439 gametophyte-biased genes, resulting in elevated π_s and π_n.
440
441 Expression level and breadth may have a stronger effect on selection efficacy than antagonistic
442 pleiotropy across life stages. Our results showed higher expression level decreased diversity and
443 the proportions of effectively neutral mutations (Figure 3, Figure S7). Gametophyte- and
444 sporophyte-biased genes likely have more specialized functions and reduced expression breadth
445 than unbiased genes, particularly pollen-biased genes (Honys and Twell 2003), which could
446 explain their relaxed constraint. However, limited tissue sampling in *R. hastatulus* (pollen and
447 leaf from a specific developmental stage) restricted our assessment of expression breadth and
448 may have affected functional enrichment (Table S5). Nonetheless, studies in *Arabidopsis*
449 *thaliana* have indicated that pollen expression is highly distinct from that of sporophytic tissues,
450 suggesting that pollen-biased genes are likely strongly enriched for gametophytic functions
451 (Rutley and Twell 2015). The lack of female gametophytic tissue (i.e., ovules) in *R. hastatulus*
452 limits our ability to detect potential sexual conflict within the gametophytic stage and between
453 sporophytes and female gametophytes. Additionally, sperm competition can contribute to the
454 maintenance of genetic variation (Clark 2002; Dapper and Wade 2016). Sperm and vegetative
455 cells in pollen might experience different selection pressures, which could explain the patterns of
456 DFE in *R. hastatulus* (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et al.
457 2022). In contrast, *C. purpureus* expression data included many cell types, the more symmetric

458 expression bias-diversity relationship (Figure 1) may reflect smaller differences in expression
459 breadth between gametophytes and sporophytes. Future studies should benefit from a wider and
460 more complete sampling of tissues from different plant species, e.g., including roots in
461 angiosperms like *R. hastatulus*.
462
463 Despite different sampling strategies, genome properties, and contrasting predominance of the
464 gametophytic stage, both species showed elevated diversity in highly gametophyte-biased genes.
465 The whole-genome resequencing data of *C. purpureus* came from geographically diverse isolates,
466 while the whole-genome sequencing data of *R. hastatulus* were from a single population (see
467 Material and Methods). Although unlikely to drive a genome-wide pattern, spatially variable
468 selection in *C. purpureus* may maintain different locally adapted alleles especially in the long-
469 lived gametophytes, increasing π and Tajima's *D*. The large non-recombining regions in *R.*
470 *hastatulus* affects gene density and thus patterns of genetic diversity (Rifkin et al. 2022). In
471 contrast, bryophytes like *C. purpureus* show higher recombination rates and more uniform gene
472 density across the genome (Gaut et al. 2007; Bowman et al. 2017; Lang et al. 2018; Healey et al.
473 2023). The consistent patterns across species support relaxed purifying selection as the most
474 plausible explanation. In contrast to antagonistic pleiotropy, the prevalence of synergistic
475 pleiotropy between life stages may explain the stronger signals of purifying selection on unbiased
476 genes, i.e., deleterious mutations in these genes are selected against in both life stages. This
477 concordant selection could dominate genome-wide patterns of diversity, and is consistent with
478 breeding studies in which selection on pollen can have synergistic pleiotropic effects on
479 sporophytic fitness (Hormaza and Herrero 1992).
480

481 *Genome-wide scan for balancing selection*

482 The maintenance of genetic variation through balancing selection is a central to evolutionary
483 biology, yet detecting balancing selection genome-wide remains challenging (Ruzicka et al.
484 2025). Traditional methods rely on the HKA test (Hudson et al. 1987) and summary statistics like
485 Tajima's *D*, which may lack power compared to composite likelihood ratio (CLR) based test
486 (Bitarello et al. 2023). We used the CLR-based *B* statistics, which adjusts window size based on
487 the data and can integrate multiple signatures including an excess of common alleles and
488 increased density of polymorphisms (Cheng and DeGiorgio 2019, 2020, 2022). However, even
489 with improved performance, *B* statistics have limited power when used to detect a single
490 signature of selection and determining appropriate window sizes is challenging without a genetic
491 map. Genome-wide assessments of selection are affected by confounding factors such as
492 demographic history and variation in recombination rates. Limited knowledge of the functional
493 annotation in the study system further complicates the interpretation of candidate sites under
494 selection. Detection power depends on the underlying selection model; for example, Flintham et
495 al. (2025) showed that balancing selection driven by sexual conflict is difficult to detect in
496 genomics, as its polymorphism signature is transient under polygenic selection. Similar
497 challenges are expected when detecting balancing selection due to intralocus conflict between life
498 stages.

499

500 Although we did not find a genome-wide signal of intralocus conflict between life stages, we
501 performed functional enrichment of candidate genes under balancing selection to test if they are
502 subject to conflict. We identified 499 and 964 genes under balancing selection in *R. hastatulus*
503 and *C. purpureus*, respectively. Similar approaches should be informative to test whether
504 sexually antagonistic SNPs are under balancing selection (Ruzicka et al. 2019). Our results show

505 over 30% of candidate genes for balancing selection (138 genes) in *R. hastatulus* are located
506 within inversion polymorphisms, consistent with findings from humans (Giner-Delgado et al.
507 2019). Since recombination is suppressed in inversion heterozygotes, many of these 138 genes
508 may show signals of balancing selection due to linkage disequilibrium rather than being direct
509 targets of selection. This linkage effect makes it difficult to distinguish focal genes under
510 balancing selection from nearby unselected genes, which potentially affects our functional
511 enrichment results (Table S12, S13). Similarly, we did not observe an enrichment of
512 gametophyte-biased genes among those under balancing selection. However, in *C. purpureus*, we
513 identified two gametophyte-biased genes that show signals of balancing selection; studies of such
514 genes could shed light on the nature of any intergenerational antagonistic pleiotropy.
515
516 *Gene expression profiles in species with biphasic life cycles*
517 The differentially expressed genes between life stages and their functional enrichment suggest
518 distinct ecological and reproductive roles for gametophytes and sporophytes. In *R. hastatulus*, the
519 smaller overlap and greater differentiation in gene expression between life stages (Table 1) may
520 reflect the evolutionary reduction of the gametophytic stage in angiosperms, where gametophytes
521 become short-lived, highly specialized, and structurally reduced compared to those in bryophytes
522 (Szövényi et al. 2011). However, differences in RNA sequencing technologies, sample sizes and
523 gene annotations methods between studies limit direct comparisons. Broader comparative
524 analyses across major land plant lineages, such as gymnosperms (Cervantes 2023), using
525 standardized methods would help clarify how life stage complexity and duration affect the extent
526 of overlapping and differential gene expression.
527

528 **Methods**

529 *Plant materials*

530 For *R. hastatulus*, we collected seeds from a population in Rosebud, Texas, US (Pickup and
531 Barrett 2013) and planted seedlings in a glasshouse at Earth Science Center, University of
532 Toronto. We collected pollen grains from the male F_1 plants for RNA isolation based on a
533 modified protocol from Lu (2011). We collected male leaf tissues for RNA isolation and female
534 leaf tissues for DNA isolation using F_1 plants, see supplementary methods for details. We used
535 the standard protocols of Qiagen Plant Mini kit and Sigma Aldrich Spectrum Plant Total RNA
536 Kits for DNA and RNA isolation, respectively. All libraries were prepared and sequenced at the
537 Genome Quebec Innovation Centre in Montréal, Canada. The DNA samples were sequenced at a
538 depth of 10-15X. The median read numbers for DNA and RNA sequences are 72,907,022 (range:
539 64,585,743 - 88,888,320) and 23,412,838 (range: 18,618,763 - 58,151,564), respectively (Table
540 S1, S2). Our final sample sizes were 20 female leaf DNA, 77 male leaf RNA, and 75 pollen
541 RNA.

542

543 For *C. purpureus*, gametophyte DNA sequences including 18 samples at 6 locations are available
544 from Carey et al. (2021c). All RNA samples were taken from juvenile (protonema) and mature
545 (gametophore) tissues at the gametophyte stage. Gametophyte RNA sequences from 3 male-
546 female sibling pairs at 2 developmental stages (each with 3 biological replicates) are available
547 from Carey et al. (2021a). We generated sporophyte tissues for RNA sequencing using the same
548 methods and at the same time with Carey et al. (2021a). The sporophyte RNA sequences
549 included 26 samples from 6 crosses, the number of replicates per cross is between 2 and 10
550 (Table S3). We removed one sporophyte RNA sample from our analyses because it had low

551 coverage. The median read number for the sporophyte RNA sequences is 52,038,814 (range:
552 33,030,143 - 64,278,575) (Table S3).
553
554 *RNA Seq data and expression analysis*
555 For *R. hastatulus*, we used the genome assembly and genome annotation from Rifkin et al. (2022)
556 and assessed the quality of RNA reads using FastQC (Andrews 2010). In *C. purpureus*, we used
557 a hybrid genome consisting of the male R40 genome assembly (autosomes, the V sex
558 chromosome) and the U sex chromosome from the female GG1 genome assembly (Carey et al.
559 2021a). We trimmed adaptor sequences and filtered out low-quality reads using trimmomatic for
560 all RNA samples (Bolger et al. 2014), and trimmed polyG sequences using fastp for the
561 sporophyte RNA samples (Chen 2023). We used FastQC to assess the quality of RNA reads
562 before and after filtering (Andrews 2010). In *C. purpureus*, we excluded four sporophyte samples
563 for being an outlier in PCA from the majority of sporophyte samples; our final sample size for
564 sporophyte RNA samples is 21.
565
566 In both species, we mapped the RNA reads to the genome assembly using STAR in two pass
567 mode (Dobin et al. 2013, v2.7.6a). We sorted the alignment SAM files and added read groups
568 using PicardTools (http://broadinstitute.github.io/picard/). We generated raw read counts for each
569 gene using featureCount (Liao et al. 2014) and then normalized the raw read counts by library
570 depth using DESeq2 (Love et al. 2014). In the expression analysis of *C. purpureus*, we included
571 both male and female gametophyte samples for autosomal genes and only included female or
572 male gametophyte samples for U and V genes, respectively. A gene was identified as expressed if
573 the mean normalized read count for either life stages was > 5; we only kept expressed genes in
574 the subsequent analyses. We calculated expression ratios between life stages for each gene by

575 dividing the mean normalized read counts at each life stage. The criteria for differentially
576 expressed genes were Benjamini-Hochberg adjusted *p*-value < 0.1, corresponding to a false
577 discovery rate of 0.1 and FC>2 (Benjamini and Hochberg 1995), allowing for a sensitive
578 characterization of biological signals and a reasonable control for false positives. A gene was
579 considered tissue-specific if: a) it was in the 10% or 90% quantiles of expression ratio among all
580 significantly differentially expressed genes (adjusted *p*-value < 0.1), and b) it had an absolute
581 value of log2FoldChange > 2. We performed GO enrichment for biased and specific genes using
582 topGO, with expressed genes from each species serving as the gene universe for comparison.
583 (Alexa and Rahnenfuhrer 2023). We divided the genes into four bins based on quantiles of
584 log2FoldChange for gametophyte-biased (log2FoldChange > 0) and sporophyte-biased
585 (log2FoldChange > 0) genes separately. We defined unbiased genes as the middle two bins with
586 lowest absolute value of log2FoldChange as well as an adjusted *p*-value ≥ 0.1. We divided the
587 expressed genes into four or ten quantiles of expression levels and used the same cutoff for later
588 analyses.
589
590 *DNA Seq data and variant calling*
591 For *R. hastatulus*, we assessed the quality of DNA raw reads using FastQC (Andrews 2010) and
592 mapped the reads to the genome assembly using bwa-mem2 (Vasimuddin et al. 2019). We sorted
593 the BAM files using SAMtools (Danecek et al. 2021), added read groups, and removed PCR
594 duplicates using PicardTools (http://broadinstitute.github.io/picard/). We called SNPs jointly on
595 all samples using freebayes (Garrison and Marth 2012) with invariant sites saved in the output (--
596 report-monomorphic). The variant and invariant sites were filtered separately using VCFtools
597 (Danecek et al. 2011), and concatenated together using BCFtools (Danecek et al. 2021). For both
598 variant and invariant sites we removed sites with a proportion of missing data > 0.8 or with a

599 mean read depth per site <10 or >40. For variant sites, we only kept bi-allelic SNPs with
600 genotype quality > 30.
601
602 For *C. purpureus*, we removed adaptor sequences using trimmomatic (LEADING:3
603 TRAILING:3 SLIDINGWINDOW:10:30 MINLEN:40) (Bolger et al. 2014), and performed
604 quality control using FastQC before and after trimming (Andrews 2010). We masked the V sex
605 chromosome when mapping female samples and masked the U sex chromosome when mapping
606 male samples using BEDTools maskfasta (Quinlan and Hall 2010). The read mapping and BAM
607 file processing methods are the same as described above in *R. hastatulus*. We called SNPs on all
608 samples jointly using BCFtools mpileup with a ploidy of 1 and the -B option (Danecek et al.
609 2021). We only kept the VCF for autosomal genes for the subsequent analysis, as the U and V
610 sex chromosomes have low recombination (Carey et al. 2021a, 2021c). We performed VCF
611 filtering using BCFtools (Danecek et al. 2021). We removed sites with a combined read depth < 5
612 and filtered variant and invariant sites separately. For variant sites, we kept bi-allelic SNPs and
613 removed sites with low quality or mapping score (QUAL<30 && MQ<30). Finally, we removed
614 sites with a proportion of missing data > 0.8 for both invariant and variant sites.
615
616 *Diversity statistics*
617 In both species we identified 0-fold and 4-fold degenerate sites in the genome using the script
618 codingSiteTypes.py
619 (https://github.com/simonhmartin/genomics_general/blob/master/codingSiteTypes.py, accessed
620 in 2020). We intersected the list of 0-fold and 4-fold sites with the filtered VCFs to generate 0-
621 fold and 4-fold VCFs for population genetic analyses. The numbers of sites (variant and
622 invariant) in each VCF are 11,432,804 (0-fold) and 2,891,775 (4-fold) in *R. hastatulus*;

623 17,394,329 (0-fold) and 4,759,819 (4-fold) in *C. purpureus*. We used pixy to calculate the 0-fold
624 and 4-fold average per site nucleotide diversity for each gene across the genome (Korunes and
625 Samuk 2021, v1.2.6.beta1). Only genes with a minimum of 50 sites for both 4-fold and 0-fold
626 nucleotide diversity were kept in the subsequent analyses.
627 We used the script popgenWindows.py
628 (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py, accessed
629 in 2020) to calculate Tajima's *D* for each gene based on the variant-only VCFs. We calculated
630 the weighted average nucleotide diversity for each bin based on the number of sites for each gene
631 (see supplementary methods). We bootstrapped the genes in each bin 1000 times to generate 95%
632 confidence intervals. We calculated the weighted average nucleotide diversity (mean and 95%
633 CI) in the same way for gametophyte-specific, sporophyte-specific and unbiased genes. We
634 removed zeros for π_s and π_n to calculate the π_n/π_s ratio for each gene.
635
636 All DNA samples in *R. hastatulus* are female, this reduces the bias in nucleotide diversity
637 estimates due to the divergence between the X and Y sex chromosomes in males. In *C.*
638 *purpureus*, nucleotide diversity was estimated from 16 DNA samples (8 males and 8 females)
639 from 5 locations, see sample information from (2021a).
640
641 *Balancing selection scan*
642 We used BalLeRMix+ with default settings to perform genomic scans for balancing selection
643 based on B_2 statistics, incorporating derived allele frequency, polymorphism and divergence
644 information (Cheng and DeGiorgio 2019, 2020, 2022). We used an outgroup to infer the
645 ancestral state for both species (supplementary methods). A high CLR indicates strong signal for
646 balancing selection. In *R. hastatulus*, the scans were performed on each chromosome using the

647 genetic positions. In *C. purpureus*, the scans were performed in 1 Mb sliding windows based on
648 the physical positions on each chromosome, using a default recombination rate of 1 cM/Mb. The
649 threshold for signature of balancing selection was set as the 5% quantile in the non-zero values of
650 composite likelihood ratio. We calculated the maximum CLR per gene. We identified the genes
651 that had at least one site with strong signal of balancing selection and also with evidence of
652 expression in this study. We performed GO enrichment for candidate balancing selection genes
653 using topGO, with expressed genes as the gene universe for comparison (Alexa and
654 Rahnenfuhrer 2023). For candidate balancing selection genes with gametophyte bias, we used
655 gametophyte-biased genes as the comparison to test the functional enrichment for balancing
656 selection.
657
658 **Data availability**
659 Raw reads of *R. hastatulus* DNA and RNA Seq are available in NCBI SRA database under
660 BioProject PRJNA744278 (to be released on Aug 1, 2025 or upon publication). Raw reads of *C.*
661 *purpureus* sporophyte RNA Seq are available through JGI genome portal, see project IDs in
662 supplementary Table S3. Scripts used in this study are available at
663 https://github.com/imengyuan/lifestage_conflict.
664
665 **Acknowledgements**
666 We thank Bill Cole and Thomas Gludovacz for help with plant maintenance, Andrew Clemens
667 for assistance with *C. purpureus* sporophyte cultivation, and Tyler Kent for help with
668 interpolating recombination rate for *R. hastatulus*. This work was supported by NSERC
669 Discovery Grants awarded to JRS, SIW and SCHB, and NSF and NASA awards to SFM. The *C.*
670 *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.

675 **References**

676 Alexa A, Rahnenfuhrer J. 2023. topGO: Enrichment analysis for gene ontology [Internet].
677 Available from: https://bioconductor.org/packages/topGO/
678 Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data [Internet].
679 Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
680 Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-
681 specific, genes show stronger purifying selection and higher rates of positive selection
682 than sporophytic genes in *Capsella grandiflora*. Mol Biol Evol. 30(11):2475–2486.
683 doi:10.1093/molbev/mst149.
684 Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano Kenji, Takano T, Yano Kentaro,
685 Matsuoka M. 2015. De novo transcriptome assembly of a fern, *Lygodium japonicum,* and
686 a web resource database, Ljtrans DB. Plant Cell Physiol. 56(1):e5–e5.
687 doi:10.1093/pcp/pcu184.
688 Beaudry FE, Rifkin JL, Barrett SC, Wright SI. 2020. Evolutionary genomics of plant
689 gametophytic selection. Plant Commun.:100115.
690 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful
691 approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 57(1):289–300.
692 Bitarello BD, Brandt DYC, Meyer D, Andrés AM. 2023. Inferring balancing selection from
693 genome-scale data. Genome Biol Evol. 15(3):evad032. doi:10.1093/gbe/evad032.
694 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
695 data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
696 Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective.
697 J Exp Bot. 60(5):1465–1478. doi:10.1093/jxb/ern355.
698 Carey SB, Jenkins Jerry, Lovell John T., Maumus Florian, Sreedasyam Avinash, Payton Adam
699 C., Shu Shengqiang, Tiley George P., Fernandez-Pozo Noe, Healey Adam, et al. 2021a.
700 Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. Sci
701 Adv. 7(27):eabh2488. doi:10.1126/sciadv.abh2488.
702 Carey SB, Kollar LM, Mcdaniel SF. 2021b. Does degeneration or genetic conflict shape gene
703 content on UV sex chromosomes? Bryophyte Divers Evol. 43(1):133–149.
704 doi:10.11646/bde.43.1.11.
705 Carey SB, Peniston JH, Payton AC, Kim M, Lipzen A, Bauer D, Lail K, Daum C, Barry K,
706 Jenkins J, et al. 2021c. Novel insights into joint estimations of demography, mutation rate,
707 and selection using UV sex chromosomes [Preprint]. bioRxiv.
708 doi:10.1101/2021.03.30.437085.
709 Cervantes S, Kesälahti R, Kumpula TA, Mattila TM, Helanterä H, Pyhäjärvi T. 2023. Strong
710 purifying selection in haploid tissue–specific genes of scots pine supports the masking
711 theory. Mol Biol Evol. 40(8):msad183. doi:10.1093/molbev/msad183.
712 Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication
713 using fastp. iMeta. 2(2):e107. doi:10.1002/imt2.107.
714 Cheng C, Kirkpatrick M. 2016. Sex-specific selection and sex-biased gene expression in humans
715 and flies. Nachman MW, editor. PLOS Genet. 12(9):e1006170.
716 doi:10.1371/journal.pgen.1006170.
717 Cheng X, DeGiorgio M. 2019. Detection of shared balancing selection in the absence of trans-
718 species polymorphism. Mol Biol Evol. 36(1):177–199. doi:10.1093/molbev/msy202.
719 Cheng X, DeGiorgio M. 2020. Flexible mixture model approaches that accommodate footprint
720 size variability for robust detection of balancing selection. Mol Biol Evol. 37(11):3267–
721 3291. doi:10.1093/molbev/msaa134.

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. doi:10.1093/bioinformatics/btab720.

Clark AG. 2002. Sperm competition and the maintenance of polymorphism. Heredity. 88(2):148–153. doi:10.1038/sj.hdy.6800019.

Crow JF, Kimura M. 1965. Evolution in sexual and asexual populations. Am Nat. 99(909):439–450. doi:10.1086/282389.

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. doi:10.1093/bioinformatics/btr330.

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. doi:10.1093/gigascience/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. doi:10.1111/evo.12848.

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. doi:10.1093/bioinformatics/bts635.

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. doi:10.1371/journal.pgen.1006130.

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. doi:10.1086/663164.

Fishman L, Saunders A. 2008. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. Science. 322(5907):1559–1562. doi:10.1126/science.1161406.

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. doi:10.1093/evlett/qrae059.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing [Preprint]. arXiv:1207.3907 [q-bio.GN]. doi:10.48550/arXiv.1207.3907.

Gerstein AC, Otto SP. 2009. Ploidy and the causes of genomic evolution. J Hered. 100(5):571–581. doi:10.1093/jhered/esp057.

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. doi:10.1038/s41467-019-12173-x.

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. doi:10.1093/molbev/mst226.

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. doi:10.1093/molbev/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. doi:10.1098/rstb.2005.1794.

769 Haldane JBS. 1932. The causes of evolution. Princeton (NJ): Princeton University Press.
770 Harrison MC, Mallon EB, Twell D, Hammond RL. 2019. Deleterious mutation accumulation in
771 *Arabidopsis thaliana* pollen genes: a role for a recent relaxation of selection. Genome
772 Biol Evol. 11(7):1939–1951. doi:10.1093/gbe/evz127.
773 Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant
774 Physiol. 132(2):640–652. doi:10.1104/pp.103.020925.
775 Hormaza JI, Herrero M. 1992. Pollen selection. Theoret Appl Genetics. 83(6–7):663–672.
776 Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on
777 nucleotide data. Genetics. 116(1):153–159. doi:10.1093/genetics/116.1.153.
778 Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. 2014. Exposing synonymous
779 mutations. Trends Genet. 30(7):308–321. doi:10.1016/j.tig.2014.04.006.
780 Immler S, Arnqvist G, Otto SP. 2012. Ploidally antagonistic selection maintains stable genetic
781 polymorphism: ploidally antagonistic selection. Evolution. 66(1):55–65.
782 doi:10.1111/j.1558-5646.2011.01399.x.
783 Johnson MG, Shaw AJ. 2016. The effects of quantitative fecundity in the haploid stage on
784 reproductive success and diploid fitness in the aquatic peat moss *Sphagnum*
785 *macrophyllum*. Heredity. 116(6):523–530. doi:10.1038/hdy.2016.13.
786 Kasimatis KR, Nelson TC, Phillips PC. 2017. Genomic signatures of sexual conflict. J Hered.
787 108(7):780–790. doi:10.1093/jhered/esx080.
788 Kidwell JF, Cleggi MT, Stewart FM. 1976. Regions of stable equilibria for models of differential
789 selection in the two sexes under random mating. Genetics. 85(1):171–183.
790 Klodová B, Fíla J. 2021. A decade of pollen phosphoproteomics. Int J Mol Sci. 22(22):12212.
791 doi:10.3390/ijms222212212.
792 Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? Nature. 351(6324):314–
793 315. doi:10.1038/351314a0.
794 Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence
795 in the presence of missing data. Mol Ecol Resour. 21(4):1359–1368. doi:10.1111/1755-
796 0998.13326.
797 Kruglyak L, Beyer A, Bloom JS, Grossbach J, Lieberman TD, Mancuso CP, Rich MS, Sherlock
798 G, Kaplan CD. 2023. Insufficient evidence for non-neutrality of synonymous mutations.
799 Nature. 616(7957):E8–E9. doi:10.1038/s41586-023-05865-4.
800 Lankinen A, Kiboi S. 2007. Pollen donor identity affects timing of stigma receptivity in *Collinsia*
801 *heterophylla* (Plantaginaceae): a sexual conflict during pollen competition? Am Nat.
802 170(6):854–863. doi:10.1086/522839.
803 Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at
804 synonymous sites in *D. melanogaster*. PLOS Genet. 9(5):e1003527.
805 doi:10.1371/journal.pgen.1003527.
806 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for
807 assigning sequence reads to genomic features. Bioinforma Oxf Engl. 30(7):923–930.
808 doi:10.1093/bioinformatics/btt656.
809 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
810 RNA-seq data with DESeq2. Genome Biol. 15(12):550. doi:10.1186/s13059-014-0550-8.
811 Lu Y. 2011. RNA isolation from *Arabidopsis* pollen grains. Bio-Protoc. 1(9):e67.
812 doi:10.21769/BioProtoc.67.
813 Mank JE. 2017. Population genetics of sexual conflict in the genomic era. Nat Rev Genet.
814 18(12):721–730. doi:10.1038/nrg.2017.83.
815 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. doi:10.1016/j.gde.2023.102129.

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. doi:10.1016/j.molp.2015.12.002.

Peters MAE, Weis AE. 2018. Selection for pollen competitive ability in mixed-mating systems. Evolution. 72(11):2513–2536. doi:10.1111/evo.13597.

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. doi:10.1002/ece3.465.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 12(1):32–42. doi:10.1038/nrg2899.

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. doi:10.1093/bioinformatics/btq033.

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. doi:10.1093/molbev/msaa271.

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. doi:10.1098/rstb.2021.0226.

Rutley N, Twell D. 2015. A decade of pollen transcriptomics. Plant Reprod. 28(2):73–89. doi:10.1007/s00497-015-0261-7.

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 [Preprint]. bioRxiv. doi:10.1101/2025.02.12.637871.

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. doi:10.1093/molbev/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. doi:10.1038/s41559-019-1041-9.

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. doi:10.1038/s41586-022-04823-w.

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. doi:10.1098/rspb.2021.0119.

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. doi:10.3389/fpls.2018.01450.

863 Slotte T, Bataillon T, Hansen TT, St. Onge K, Wright SI, Schierup MH. 2011. Genomic
864 determinants of protein evolution and polymorphism in *Arabidopsis*. Genome Biol Evol.
865 3:1210–1219. doi:10.1093/gbe/evr094.
866 Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ. 2011. Generation-biased gene expression
867 in a bryophyte model system. Mol Biol Evol. 28(1):803–812.
868 doi:10.1093/molbev/msq254.
869 Szövényi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. 2013. Selection is no more
870 efficient in haploid than in diploid life stages of an angiosperm and a moss. Mol Biol
871 Evol. 30(8):1929–1939. doi:10.1093/molbev/mst095.
872 Travers SE, Mazer SJ. 2001. Trade-offs between male and female reproduction associated with
873 allozyme variation in phosphoglucoisomerase in an annual plant (*Clarkia unguiculata*:
874 Onagraceae). Evolution. 55(12):2421–2428. doi:10.1111/j.0014-3820.2001.tb00757.x.
875 Urrutia AO, Hurst LD. 2003. The signature of selection mediated by expression on human genes.
876 Genome Res. 13(10):2260–2264. doi:10.1101/gr.641103.
877 Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-
878 MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed
879 Processing Symposium (IPDPS). p. 314–324.
880 Walbot V, Evans MMS. 2003. Unique features of the plant life cycle and their consequences. Nat
881 Rev Genet. 4(5):369–379. doi:10.1038/nrg1064.
882 Wright AE, Fumagalli M, Cooney CR, Bloch NI, Vieira FG, Buechel SD, Kolm N, Mank JE.
883 2018. Male-biased gene expression resolves sexual conflict through the evolution of sex-
884 specific genetic architecture. Evol Lett. 2(2):52–61.
885 Zhang J, Qian W. 2025. Functional synonymous mutations and their evolutionary consequences.
886 Nat Rev Genet.:1–16. doi:10.1038/s41576-025-00850-1.
887 Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. Nat Rev Genet.
888 16(7):409–420. doi:10.1038/nrg3950.
889 Zhou Z, Dang Y, Zhou M, Li L, Yu C, Fu J, Chen S, Liu Y. 2016. Codon usage is an important
890 determinant of gene expression levels largely through its effects on transcription. Proc
891 Natl Acad Sci. 113(41):E6117–E6125. doi:10.1073/pnas.1606724113.