Testing for the genomic footprint of conflict between life stages in an angiosperm and a 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 (2n) 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 are 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* are 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, no empirical studies have investigated this problem. Our results revealed genome-wide patterns more consistent with synergistic pleiotropy, with gametophyte-biased genes mostly likely to be under relaxed purifying selection in both 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). 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

72 selection on gametophytes (gametophytic selection) affect sporophytes is a key unresolved
73 question in plant evolutionary genetics.
74
75 The evolutionary consequences of gametophytic selection have been examined both theoretically
76 and empirically (Haldane 1932; Otto et al. 2015; Immler and Otto 2018; summarized in Immler
77 2019, Beaudry et al. 2020). The haploid nature of the gametophytic phase is predicted to increase
78 the efficacy of selection due to the lack of dominance (i.e., the masking hypothesis), allowing for
79 more efficient removal of recessive deleterious mutations (i.e., purging) and fixation of recessive
80 beneficial mutations (Crow and Kimura 1965; Kondrashov and Crow 1991; Gerstein and Otto
81 2009). Consistent with this prediction, some studies have found stronger purifying and positive
82 selection in gametophyte-specific than sporophyte-specific genes (Arunkumar et al. 2013;
83 Cervantes 2023; but see Harrison et al. 2017). Stronger selection on gametophyte-expressed
84 genes can also be driven by gametophytic competition (sexual selection) in angiosperms
85 (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et al. 2022). Additionally,
86 gene expression breadth can be a stronger predictor of selection efficacy than the masking effect
87 in angiosperms and bryophytes (Szövényi et al. 2013).
88
89 The common genome shared between life stages provides an opportunity to test how
90 gametophytic selection affects the evolution of the sporophyte stage due to the pleiotropy
91 between life stages. Plants share a substantial overlap in gene expression between life stages
92 based on both microarray and mRNA sequencing studies (summarized in Beaudry et al. 2020). In
93 ferns and bryophytes where the gametophytes are photosynthetic and exhibit indeterminate
94 growth, gene expression overlap may exceed that in angiosperms, e.g., estimates are ~60% in the
95 angiosperm *Arabidopsis thaliana* (Walbot and Evans 2003; Borg et al. 2009), 97.7% in the fern

96 *Polypodium amorphum* (Sigel et al. 2018), and 85% in the bryophyte *Physcomitrium patens*
97 (Ortiz-Ramírez et al. 2016). Previous studies on gametophytic selection have compared selective
98 forces in gametophyte -*vs.*- sporophyte-specific genes to test for the effects of haploid expression
99 on the distribution of selection coefficients (Arunkumar et al. 2013; Gossmann et al. 2014;
100 Harrison et al. 2019; Gutiérrez-Valencia et al. 2021; Cervantes et al. 2023). However, the
101 patterns of genetic diversity and selection pressures of genes with overlapping expression, which
102 are more likely to be pleiotropic due to the greater potential for competing selection pressures
103 across life stages (Immler et al. 2012; Peters and Weis 2018), have not yet been specifically
104 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). Similarly, in the case of intralocus conflict between life stages, we predict stronger
128 signals of balancing selection in genes with overlapping expression between life stages,
129 particularly those with relatively similar levels of expression between the two life stages (Immler
130 et al. 2012; Peters and Weis 2018), reflected by diversity statistics such as nucleotide diversity
131 (π) and Tajima's D. In addition, we also predicted that candidate genes under balancing selection
132 would be disproportionately biased in expression between life stages rather than being unbiased.
133
134 We examined gene expression and genetic diversity patterns in the gametophytic and sporophytic
135 life stages of two plant species: the angiosperm *Rumex hastatulus* (Polygonaceae) and the
136 bryophyte *Ceratodon purpureus* (Ditrichaceae). The annual *R. hastatulus* is dioecious, obligately
137 outcrossing, and wind-pollinated with an XY sex chromosome system (Smith 1964; Rifkin et al.
138 2021). The uniovulate flowers and the large quantity of pollen commonly produced in wind-
139 pollinated angiosperms suggests high potential for gametophytic selection during pollen
140 competition (Friedman and Barrett 2009; Field et al. 2012; Immler 2019). In the dioicous moss
141 *C. purpureus*, the female and male gametophytes possess the U or V sex chromosome,
142 respectively (Carey et al. 2021a), they produce gametes that fuse to form a diploid zygote (i.e.,
143 sporophyte, unsexed); the sporophyte matures and ultimately undergoes meiosis, but its entire

144 development occurs on the female gametophyte plant, providing multiple opportunities for
145 antagonistic pleiotropy (Carey et al. 2021b; McDaniel 2023). The contrasts in the length and
146 complexity of the gametophytic stage in these two species allow us to evaluate the evolutionary
147 consequences of biphasic plant life cycles in a broader context.
148
149 We first compared gene expression in gametophytic and sporophytic tissues of *R. hastatulus* and
150 *C. purpureus,* finding a larger overlap in gene expression and a smaller proportion of biased
151 genes between life stages in *C. purpureus* than *R. hastatulus*. We then examined the population
152 genetic signals of balancing selection due to antagonistic pleiotropy between life stages by
153 examining nucleotide diversity and Tajima's *D* across varying degrees of expression bias
154 between life stages. In contrast with our predictions for genome-wide patterns of antagonistic
155 pleiotropy, our results show elevated nucleotide diversity and Tajima's *D* in highly gametophyte-
156 biased genes in both *R. hastatulus* and *C. purpureus*. We found that expression levels and breadth
157 were a stronger predictor of selection efficacy than potential competing selective pressures
158 between life stages. Lastly, we identified in both species genomic regions containing hundreds of
159 candidate genes under balancing selection using a composite likelihood ratio test (Cheng and
160 DeGiorgio 2019, 2020, 2022), and examined the biological functions and differential gene
161 expression patterns of these candidate genes.
162
163 **Results**
164 *Gene expression in gametophytes and sporophytes*
165 We compared gene expression in gametophyte and sporophyte tissues of *R. hastatulus* and *C.*
166 *purpureus*. In *R. hastatulus*, we sequenced 75 mature pollen samples and 77 male leaf samples
167 representing the gametophyte and sporophyte stages, respectively (Table S1). In *C. purpureus*,

168 we 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 showed evidence of expression. In *C. purpureus*, we identified
173 24,097 gametophyte-expressed genes, consisting of 69% of all annotated genes and 93% of genes
174 with evidence of expression. The overlap in gene expression between life stages (89% in *R.*
175 *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 (padj < 0.1, FC > 2). Among
184 significant DE genes, we filtered for genes in the top 10% quantile of expression bias, allowing
185 us to compare genes with different tissue specificities. Our gene filtration process resulted in
186 similar numbers of gametophyte- and sporophyte-specific genes: 1,711 gametophyte-specific and
187 1,711 sporophyte-specific genes in *R. hastatulus*, and 1,886 gametophyte-specific and 1,867
188 sporophyte-specific genes in *C. purpureus*. Gene ontology (GO) enrichment analyses indicated
189 very different functional enrichment for biased and specific genes between life stages (Table S5,
190 S6), highlighting distinctive expression profiles across life stages in both species. For example, in
191 *R. 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. 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%)

198

199

200 *Nucleotide diversity and expression bias between life stages*

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

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

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

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

205 direction and degree of expression bias between life stages (log2FoldChange) showed a

206 significant and positive relation, but with near-zero slopes in both species (Figure S1, S2). We

207 next divided the genes into eight bins based on the direction and degree of expression bias

208 between life stages and calculated the weighted mean of π_s and π_n for each bin, weighted by the

209 number of sites per gene. In both species, highly gametophyte-biased genes exhibited the highest

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

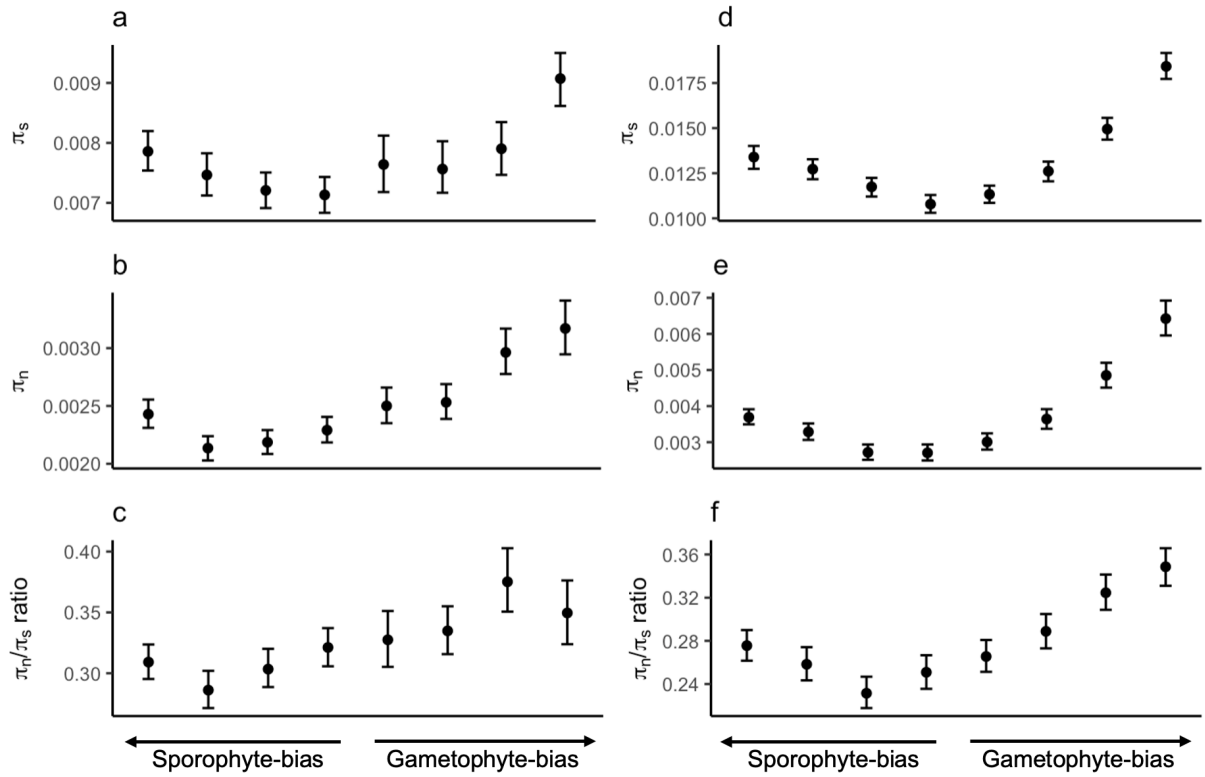
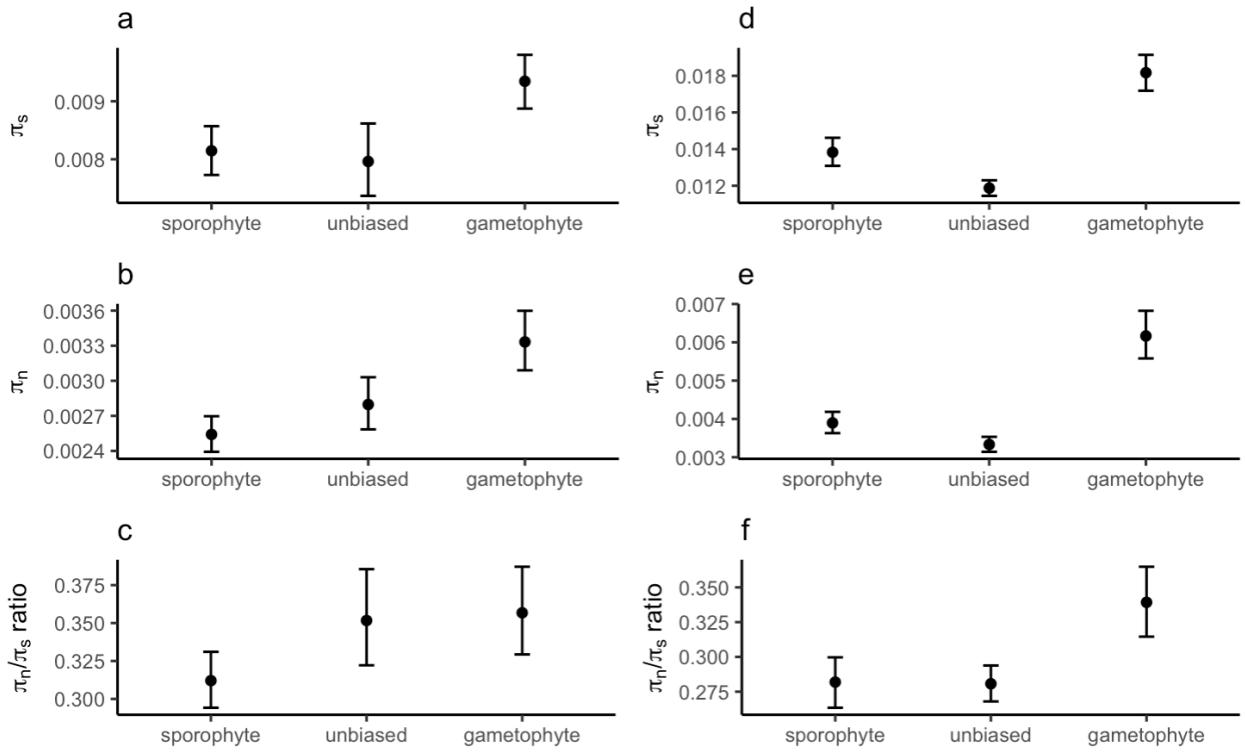


Figure 1. Effect of expression bias between life stages on nucleotide diversity in *R. hastatulus* (a-c) and *C. purpureus* (d-f). x-axis: quantiles of log2FoldChange between gametophyte and sporophyte expression. Number of genes in each bin: 1,009-1,847 for *R. hastatulus*; 1,943-2,564 for *C. purpureus*. 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

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

243 ratios are not elevated above unbiased genes in *R. hastatulus* like *C. purpureus* (Figure 2c, 2f).



244

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

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

247 1149 (s), 920 (g), 523 (unbiased) for *R. hastatulus*; 1,232 (s), 1,217 (g), 3,309 (unbiased) for *C.*

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

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

250

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

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

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

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

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

bins of expression level (baseMean across all samples) to examine how expression level affects π_s , π_n and π_n/π_s in gametophyte- and sporophyte-biased genes. As expression level increases, both π_n and π_s decrease, with a bigger decrease in π_n than π_s (Figure 3). Gametophyte-biased genes have higher π_s, π_n and π_n/π_s than sporophyte-biased genes, regardless of expression level in both species, except for the bins with overlapping confidence intervals in π_s of *R. hastatulus* (Figure 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 do not show signals of elevated diversity due to balancing selection. Instead, heterogeneity in diversity likely reflects differences in the strength of purifying selection.

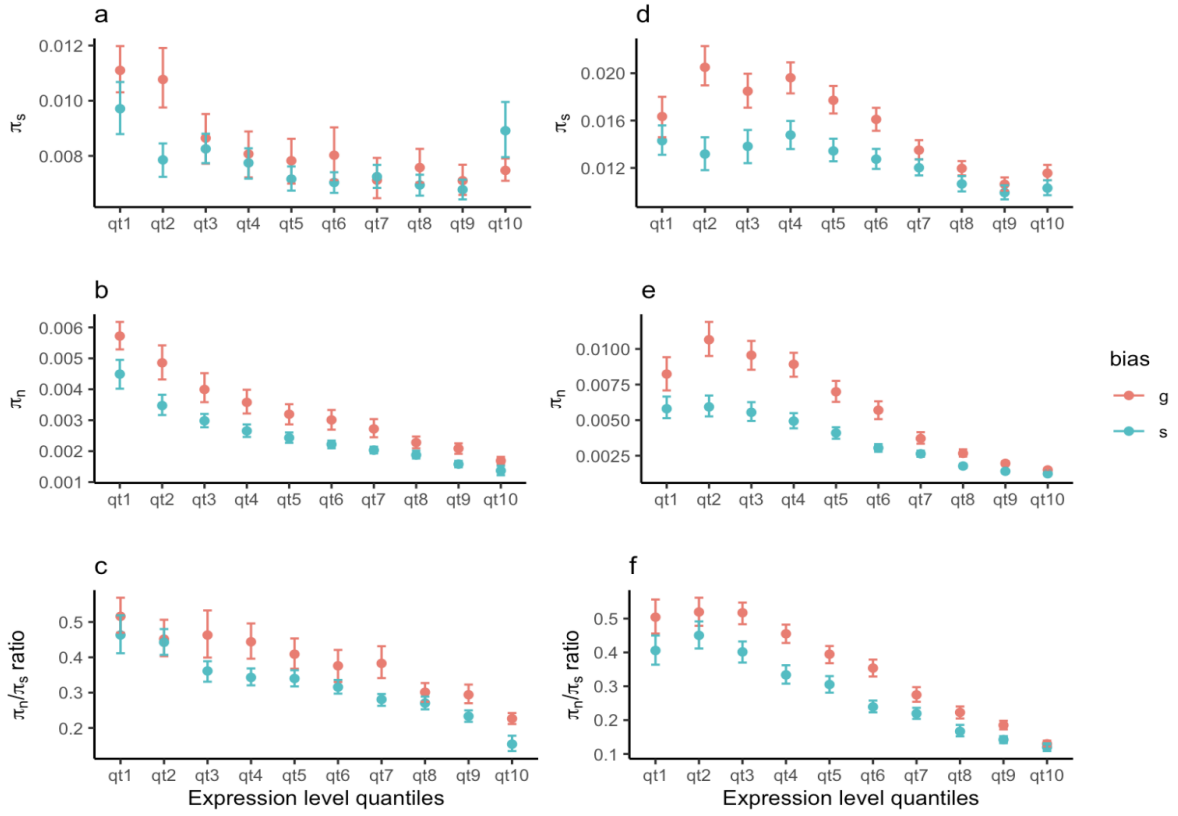
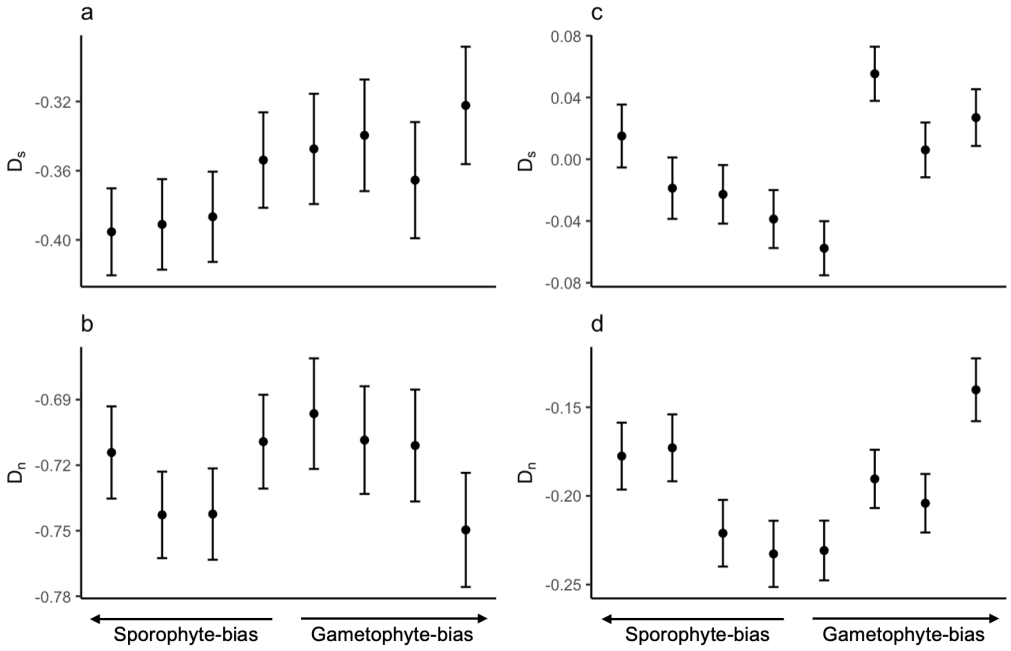


Figure 3. Effect of expression level on weighted mean nucleotide diversity in *R. hastatulus* (a-c) and *C. purpureus* (d-f). Number of genes in each bin: 686-1,464 for *R. hastatulus*; 1,034-2,328 for *C. purpureus*. The weighted mean and 95% confidence intervals are based on 500 bootstraps

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

292 levels in both species. Combined with patterns of π (Figure 1-3) our results are consistent with

293 the hypothesis that gametophyte-biased genes are under relaxed purifying selection.



294

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

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

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

298 2,138-2,534 (D_s), 2,097-2,579 (D_n) for *C. purpureus*. Error bars represent mean ± SEM across

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

300

301 *Distribution of fitness effects*

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

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

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

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

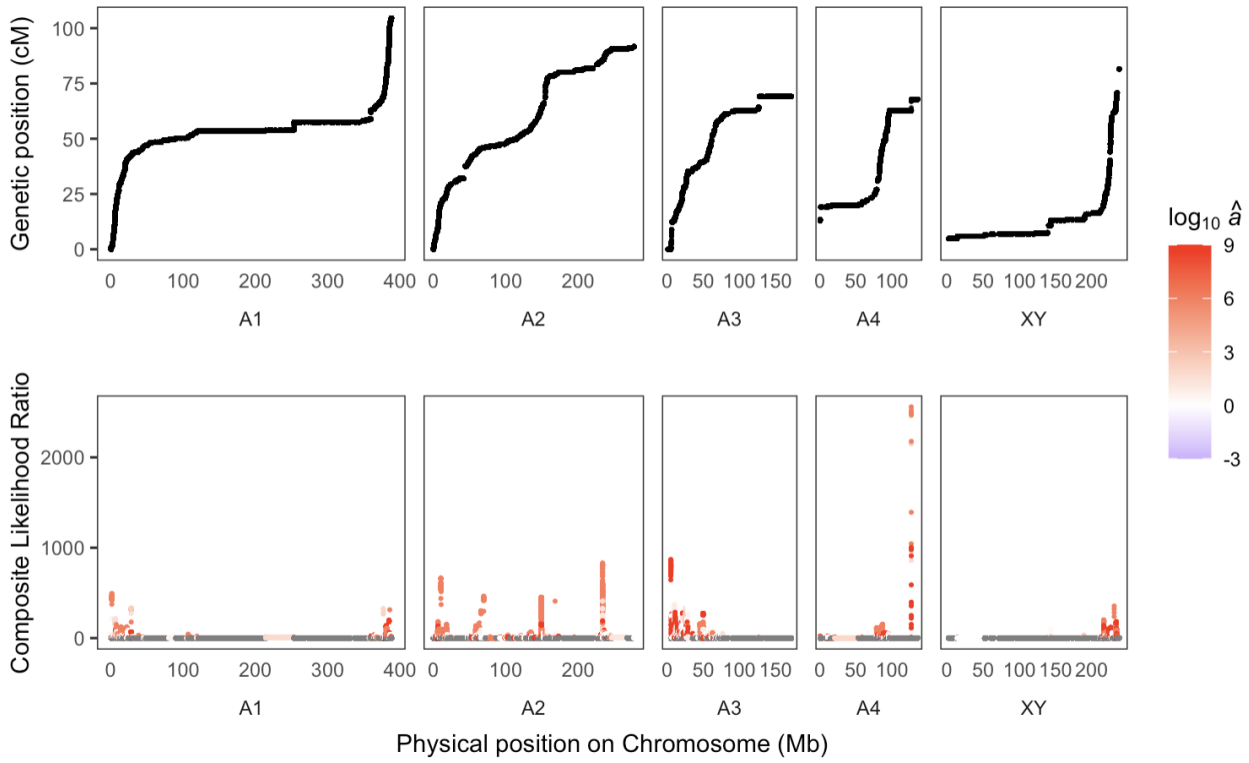
306 first compared the DFE of gametophyte-specific, sporophyte-specific and unbiased genes (Figure

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

331 (Arunkumar et al. 2013; Gossmann et al. 2014; Harrison et al. 2019; Gutiérrez-Valencia et al.
332 2021; Cervantes et al. 2023), our results in both species did not support the haploid purging
333 hypothesis. In *R. hastatulus*, gametophyte-specific genes had a higher proportion of strongly
334 deleterious mutations than sporophyte-specific genes, in *C. purpureus* the proportions were
335 similar with overlapping confidence intervals (Figure S6). In the low to intermediate expression
336 quantiles (qt1 to qt3), sporophyte-specific genes showed a slightly higher proportion of strongly
337 deleterious mutations ($N_e s > 10$) than gametophyte-specific genes in both species (Figure S7). In
338 the highest expression quantile (qt4), gametophyte-specific genes had slightly higher proportions
339 of strongly deleterious mutation ($N_e s > 10$) than sporophyte-specific genes in both species, but
340 the differences were not significant with overlapping confidence intervals (*R. hastatulus*: 67.19 -
341 70.21% in n, 65.38 - 68.66% in 2n; *C. purpureus*: 77.09 - 80.36% in n, 74.22 - 78.78 % in 2n,
342 Figure S7, Table S10).
343
344 *Genome-wide scan for balancing selection*
345 These overall patterns of diversity do not show a genome-wide signal of an enrichment of
346 balancing selection on genes with overlapping expression between the life stages; however, it is
347 possible that individual genes may still be subject to balancing selection. We therefore took an
348 alternative approach by conducting a genome-wide scan for signals of balancing selection and
349 explored whether they were enriched for genes that might be subject to antagonistic pleiotropy
350 between life stages. We applied a composite likelihood ratio test to scan the genome for
351 candidate genes under balancing selection using values of polymorphism and divergence (Cheng
352 and DeGiorgio 2019, 2020, 2022). In *R. hastatulus,* a total number of 4,945,561 sites were tested
353 in windows based on genetic positions (Figure 4). We found 79,431 sites with strong signals of
354 balancing selection across all chromosomes (CLR > 9.5), located at 449 genes, 405 of which had

355 evidence of expression. The candidate genes were only located in regions with considerable
356 recombination, based on our genetic maps (Figure 4). However, among the genes found in
357 regions with evidence of balancing selection, 138 of them were located within inversion
358 polymorphisms identified from two phased whole genome assemblies of *R. hastatulus* on
359 chromosomes A2 (63), A3 (74), and A4 (1) (Sacchi et al. 2024; Sacchi et al, in prep), consisting
360 of more than 30% of all candidate genes; whereas these inversions contained 16.7% of genes
361 across the genome. Because inversion heterozygotes are likely to experience reduced
362 recombination, these candidate genes may be subject to non-independent signals of balanced
363 polymorphism. In *C. purpureus*, the scans were performed in physical position windows as site-
364 specific recombination rates were unavailable (Figure S8). We tested 7,743,902 sites and
365 identified 339,375 sites under balancing selection (CLR > 956.2), located at 963 genes, 807 of
366 which had evidence of expression in this study. In *R. hastatulus*, candidate genes under balancing
367 selection had significantly higher π_s, and significantly lower π_n/π_s than genes that were not under
368 balancing selection (Welch Two Sample *t*-test, *p*-value = 0.0001511, 0.0003204, respectively). In
369 *C. purpureus*, there was no significant differences in π_s or π_n/π_s between genes under balancing
370 selection and gene that were not (Welch Two Sample *t*-test, *p*-value = 0.6029; Two Sample *t*-test,
371 *p*-value = 0.5956, respectively).
372
373 We performed contingency tests on the number of genes in different expression bias categories
374 (gametophyte-biased, sporophyte-biased, unbiased) for genes under balancing selection and
375 across the whole genome (Tables S11). In both species, we found no enrichment of any category
376 of expression bias under long-term balancing selection compared to the whole genome (Fisher's
377 exact test, p > 0.05). We examined the gene functions for all gametophyte-biased genes in
378 regions under balancing selection (Table S12, S13). We performed GO enrichment after

379 excluding the genes found within inversion polymorphisms in *R. hastatulus*, since the
380 suppression of recombination between inversion types can influence our ability to find functional
381 enrichments of causal loci. The GO enrichment results after excluding inversions were not
382 compelling (Table S12). In *R. hastatulus*, two genes under balancing selection had probable
383 disease resistance functions based on orthologs to *Arabidopsis thaliana* (At4g19060,
384 At4g14610). In *C. purpureus*, two genes under balancing selection were expressed in both life
385 stages with high gametophyte bias and have sexual reproduction functions (log2FoldChange > 3,
386 GO:0019953). These genes were also in the top 10% quantile of π_s.



387

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

391

392 **Discussion**

393 We combined population genomic and differential gene expression analyses to examine the
394 population genetic signals of balancing selection due to potential conflict between life stages in
395 *R. hastatulus* and *C. purpureus*. First, we showed that there is larger overlapping gene expression
396 and less differential expression in the moss *C. purpureus* than in the angiosperm *R. hastatulus,*
397 based on our sampling of tissues. If most of the signature of balancing selection is attributable to
398 antagonistic pleiotropy between life stages, we would expect a signal of greater balancing
399 selection on genes with overlapping expression between life stages. In contrast, we found
400 elevated π_s, π_n and D_n in highly gametophyte-biased genes in both species. Combined with the
401 observations that π_s, π_n and π_n/π_s all decreased as gene expression level increased, the elevated
402 diversity suggests relaxed purifying selection in highly gametophyte-biased genes rather than
403 balancing selection. In addition, based on DFE comparisons we found no evidence that
404 gametophyte-specific genes undergo stronger purifying selection than sporophyte-specific genes.
405 Lastly, we performed genome-wide scans and identified hundreds of candidate genes under
406 balancing selection in both species; however, we did not observe an enrichment of gametophyte-
407 biased genes in candidates. Below we discuss the implications of our results for the evolutionary
408 consequences of gametophytic selection in species in which the gametophytic life stage differs in
409 duration and the challenges of using genome-wide population genomics to test models of
410 conflict.

411

412 *Relaxed purifying selection rather than balancing selection*

413 Balancing selection generated by opposing selective pressures between sexes (sexual
414 antagonism) has been tested in a series of empirical studies using genetic diversity and sex-biased
415 expression (Cheng and Kirkpatrick 2016; Kasimatis et al. 2017; Mank 2017; Wright et al. 2018;

416 Sayadi et al. 2019). These studies have shown elevated diversity in weakly sex-biased genes
417 consistent with the prediction of intralocus sexual conflict. This is because highly sex-biased
418 genes likely reflect resolved conflict with little potential for ongoing antagonism, whereas fully
419 unbiased genes may be enriched in genes with highly concordant effects in males and females.
420 Similarly, we predicted higher genetic diversity with lower or intermediate expression bias due to
421 intralocus conflict between life stages. In contrast, we found higher π_s, π_n and D_n in highly
422 gametophyte-biased genes in both species (Figure 1), which may be explained by balancing
423 selection or relaxed purifying selection on nonsynonymous and synonymous sites in highly
424 gametophyte-biased genes. As highly expressed genes are under stronger selective constraint
425 than lowly expressed genes (e.g., Urrutia and Hurst 2003; Slotte et al. 2011; Zhang and Yang
426 2015), an observation of high π_s in gametophyte-biased genes with high expression level would
427 suggest balancing selection. However, π_s decreases with higher expression level in both
428 gametophyte- and sporophyte-biased genes in both species (Figure 3), and thus we conclude that
429 relaxed purifying selection on both nonsynonymous and synonymous sites is the more plausible
430 explanation. Overall, our results are not consistent with our predictions for widespread intralocus
431 conflict between the life stages. However, other processes are likely to contribute to the patterns
432 of elevated genetic variation but were not tested in this study; below we discuss the alternative
433 processes and caveats based on our dataset.
434
435 Two processes could lead to higher π_s in genes subject to relaxed purifying selection: weaker
436 background selection, or less direct selection on synonymous sites themselves, or both.
437 Background selection reduces neutral nucleotide diversity due to selection on linked deleterious
438 sites; the reduction in diversity is larger when the linked region contains more constrained sites
439 (Elyashiv et al. 2016). Highly gametophyte-biased genes could be subject to weaker background

440 selection, resulting in elevated π_s and π_n. Synonymous mutations are generally assumed to be
441 neutral as they do not affect amino acid sequences of the encoded protein (King and Jukes 1969).
442 For example, DFE analyses often use synonymous (4-fold) sites as a neutral control to compare
443 the site frequency spectra with nonsynonymous (0-fold) selected sites (Keightley and Eyre-
444 Walker 2007). However, studies in fruit flies (*Drosophila melanogaster*) and yeast
445 (*Saccharomyces cerevisiae*) found that 22% and three quarters of examined synonymous
446 mutations are strongly deleterious, respectively (Lawrie et al. 2013; Shen et al. 2022; but see
447 Kruglyak et al. 2023). One explanation for selection on synonymous mutations is due to codon
448 usage bias, where synonymous codons that lead to more efficient translation are favored (Plotkin
449 and Kudla 2011; Hunt et al. 2014). Codon usage bias is also found to affect transcription and
450 gene expression levels (Zhou et al. 2016). A study in the liverwort *Marchantia polymorpha*
451 found relaxed purifying selection in sporophyte-biased genes and evidence for lower codon usage
452 bias in these genes (Shen et al. 2024). In our case, both synonymous and nonsynonymous sites
453 may be under constraint, and the level of constraint may be lower in highly gametophyte-biased
454 genes, resulting in elevated π_s and π_n.
455
456 Expression level and expression breadth across multiple tissues may have a stronger effect on
457 selection efficacy than the antagonistic pleiotropy across life stages. Many studies have found
458 expression level as a strong factor in predicting selection efficacy (e.g., Urrutia and Hurst 2003;
459 Slotte et al. 2011; Zhang and Yang 2015). Consistent with these findings, our results show that π_n
460 is lower than π_s across expression levels (Figure 3) and there is a decrease in π_n/π_s (Figure 3) and
461 the proportions of effectively neutral mutations (Figure S7) with higher expression level. If
462 gametophyte- and sporophyte-biased genes are likely to have more specialized functions than
463 unbiased genes, the reduced expression breadth in highly gametophyte-biased genes could

464 potentially explain why they are under relaxed purifying selection than unbiased genes. In
465 contrast with antagonistic pleiotropy, the prevalence of synergistic pleiotropy between life stages
466 could explain why we observe signals of stronger purifying selection on unbiased genes (e.g.,
467 deleterious mutations in unbiased genes are under purifying selection in both life stages), and this
468 may predominate in affecting patterns of genome-wide diversity over signals of antagonistic
469 pleiotropy elevating neutral diversity in these genes.
470
471 The limited sampling of tissues especially in the angiosperm *R. hastatulus* makes it challenging
472 to accurately account for expression breadth and to differentiate alternative processes that might
473 contribute to the patterns we observed. In *R. hastatulus*, we only used one tissue per life stage
474 (pollen for gametophyte and leaf for sporophyte), female gametophytic tissue (i.e., ovule) and
475 other sporophytic tissues (e.g, roots, stem, etc.) were not sampled. As pollen is a highly
476 specialized tissue, pollen-biased genes will likely have a smaller expression breadth than
477 unbiased or leaf-biased genes (e.g., Honys and Twell 2003). The lack of female gametophytic
478 tissue (i.e., ovules) in *R. hastatulus* makes it difficult to capture the potential sexual conflict
479 within the gametophytic stage (i.e., between pollen and ovule) and conflict between sporophytes
480 and female gametophytes; as a result, other potential forms of pleiotropy may have been missed.
481 Sperm competition can also contribute to the maintenance of genetic variation (Clark 2002;
482 Dapper and Wade 2016). The sperm and vegetative cells within pollen tissue might experience
483 different strengths of selection which could explain the patterns we see in the DFE results of *R.*
484 *hastatulus* (Arunkumar et al. 2013; Gossmann et al. 2014; Gutiérrez-Valencia et al. 2022). In *C.*
485 *purpureus*, we generated expression data from gametophytic and sporophytic tissues that include
486 many cell types. Why we observed more relaxed purifying selection in highly gametophyte- than
487 sporophyte-biased or unbiased genes in *C. purpureus* is less clear. The increase in π_s and π_n with

488 stronger expression bias between the life stages is more symmetrical in *C. purpureus* than *R.*
489 *hastatulus*. This result suggests there is a smaller difference in expression breadth between
490 gametophytes and sporophytes with multiple cell types in *C. purpureus* than pollen and leaf
491 tissues in *R. hastatulus.* Future studies will benefit from a wider and more complete sampling of
492 tissues from different plant species.
493
494 It is noteworthy that we found a common pattern of elevated genetic diversity in highly
495 gametophyte-biased genes in our two study species with contrasting predominance of the
496 gametophytic stage, despite the differences in data collection methods and genome properties.
497 The source population for nucleotide polymorphism data and genome properties, such as
498 recombination landscapes, differ in our two study species. The whole-genome resequencing data
499 of *C. purpureus* are generated from isolates collected from five locations across the Northern
500 Hemisphere (Carey et al. 2021c). The whole-genome sequencing data of R. *hastatulus* were
501 obtained from a single population in Texas, US (Pickup and Barrett 2013). The large non-
502 recombining regions in *R. hastatulus* affects gene density and, as a result, patterns of genetic
503 diversity across the genome (Rifkin et al. 2022). In contrast, bryophyte species have a higher
504 average recombination rate than vascular plants, the distribution of gene density is more uniform
505 across the genome (Gaut et al. 2007; Bowman et al. 2017; Lang et al. 2018; Healey et al. 2023).
506 As the genomic data from *C. purpureus* consist of populations from ecologically different
507 habitats, spatially variable selection could potentially maintain different locally adapted alleles
508 especially in the long-lived gametophytes, which would increase nucleotide diversity and
509 Tajima's *D* in gametophyte-biased than sporophyte-biased genes. Because both species show
510 similar patterns of both π_s and π_n despite differences, relaxed purifying selection in highly
511 gametophyte-biased genes is likely to be the most plausible explanation for both species.

512

513 *Genome-wide scan for balancing selection*

514 The maintenance of genetic variation is a key question in evolutionary biology, but determining
515 the genomic prevalence of balancing selection remains an unresolved issue despite intensive
516 effort. Standard methods of detecting balancing selection often use summary statistics like π and
517 Tajima's D, which could lack power compared to composite likelihood ratio (CLR) based
518 methods (summarized in Bitarello et al. 2023). We identified hundreds of genes in regions under
519 balancing selection across the genome in *R. hastatulus* (449) and *C. purpureus* (964) using a
520 CLR-based method called B statistics (Cheng and DeGiorgio 2019, 2020, 2022), within which
521 138 genes were located within inversion polymorphisms previously identified in *R. hastatulus*.
522 Note that because recombination should be suppressed in inversion heterozygotes, many of these
523 138 genes may be showing the signal of balancing selection due to linkage disequilibrium, rather
524 than being direct targets of balancing selection. Even though we did not find a genome-wide
525 pattern for balancing selection due to intralocus conflict between life stages in both species, we
526 performed functional enrichment of candidate genes under balancing selection to identify genes
527 subject to intralocus conflict between life stages. Similar approaches will be useful in sexual
528 conflict studies (Ruzicka et al. 2019), e.g., to test whether the identified sexually antagonistic
529 SNPs are under balancing selection. We identified two genes expressed in both life stages of *C.*
530 *purpureus* that show gametophyte bias and signals of balancing selection; studies of such genes
531 could shed light on the nature of any intergenerational antagonistic pleiotropy. Although we did
532 not see an enrichment of gametophyte-biased genes under balancing selection compared to the
533 whole genome, or more compelling functional enrichment, it is noteworthy that focal genes
534 under balancing selection and nearby unselected genes are indistinguishable due to linkage,
535 which affect our functional enrichment results. Challenges also remain with CLR-based methods

536 in interpreting results when a genetic map is unavailable, as information of recombination rate is
537 important in determining the window sizes of different detection methods (Bitarello et al. 2023).
538
539 Studies on balancing selection are important in helping the understanding of how polymorphisms
540 in structural variants such as inversions are maintained. Chromosomal inversions play a key role
541 in evolutionary processes such as local adaptation and speciation (Dobzhansky 1970; Kirkpatrick
542 and Barton 2006; Huang and Rieseberg 2020). Inversion polymorphisms in a population can be
543 identified through genomic studies (e.g., Sacchi et al. 2024). Studies in humans found an
544 enrichment of balancing selection signatures in inversion polymorphisms (Giner-Delgado et al.
545 2019). Similarly, our results show over 30% of candidate genes for balancing selection are
546 located within inversion polymorphisms in *R. hastatulus*. The potential predominance of
547 balancing selection signatures within inversion polymorphisms suggests a role of balancing
548 selection in maintaining inversion polymorphisms in populations (Dobzhansky 1970; Kirkpatrick
549 and Barton 2006). Future studies using genomic data from more taxa including trans-specific
550 polymorphisms will provide more evidence on the biological functions and evolutionary roles of
551 balancing selection.
552
553 *Gene expression profiles in species with biphasic life cycles*
554 The differentially expressed genes between life stages and their functional enrichment suggest
555 distinctive ecological and reproductive roles for gametophytes and sporophytes. There is a
556 smaller overlap in gene expression (87% vs 93%) and a larger proportion of differentially
557 expressed genes (79% vs 45%) between life stages in *R. hastatulus* than in *C. purpureus*. The
558 smaller overlap and greater differentiation in gene expression in the angiosperm *R. hastatulus*
559 may reflect the reduction of the gametophytic stage during land plant evolution, as gametophytes

560 become short-lived, more specialized and much more reduced in angiosperms than bryophytes.
561 Similar results were found when comparing the moss *Funaria hygrometrica* and the angiosperm
562 *Arabidopsis thaliana* (Szövényi et al. 2011). Similarly, the overlap in gene expression is larger in
563 the fern *Polypodium amorphum* (97.7%) of which the gametophytes are usually photosynthetic
564 and sometimes independent when mature compared to angiosperms (Sigel et al. 2018).
565 Furthermore, the proportions of sporophyte- and gametophyte-biased genes are more similar in
566 *C. purpureus* than in *R. hastatulus*, with slightly more biased genes towards the more dominant
567 life stage in each species respectively (Table 1). However, the RNA sequencing technologies,
568 sample sizes and the nature of gene annotations are different between these studies. Therefore,
569 comparative analyses using a wider sampling of species from all major land plant lineages, e.g.,
570 gymnosperms (Cervantes 2023), and directly comparable methods would be useful to test how
571 the varying complexity and length of life stages affects the proportion of overlapping and
572 differential gene expression between life stages. Overall, our results suggest the two plant life
573 stages undergo more concordant than discordant selection on gene expression despite differences
574 in their ploidy, genetics, development and morphology. This result is consistent with previous
575 plant-breeding studies showing selection on pollen has synergistic pleiotropic effects on the
576 fitness of sporophytic offspring, e.g., cold treatment in pollen increases cold tolerance in the
577 offspring in maize (reviewed in Hormaza and Herrero 1992). Future studies on the
578 developmental and evolutionary genetics of gametophytes and sporophytes in more plant species
579 will benefit both biotechnology and the understanding of land plant evolution.
580

581 **Methods**

582 *Plant materials*

583 For *R. hastatulus*, we collected seeds from a population in Rosebud, Texas, US (Pickup and
584 Barrett 2013) and planted seedlings in a glasshouse at Earth Science Center, University of
585 Toronto. We collected pollen grains from the male F_1 plants for RNA isolation based on a
586 modified protocol from Lu (2011). We collected leaf tissues from male and female F_1 plants for
587 DNA and RNA isolation. We used the standard protocols of Qiagen Plant Mini kit and Sigma
588 Aldrich Spectrum Plant Total RNA Kits for DNA and RNA isolation, respectively. All libraries
589 were prepared and sequenced at the Genome Quebec Innovation Centre in Montréal, Canada.
590 The DNA samples were sequenced at a depth of 10-15X. The median read numbers for DNA and
591 RNA sequences are 72,907,022 (range: 64,585,743 - 88,888,320) and 23,412,838 (range:
592 18,618,763 - 58,151,564), respectively (Table S1, S2).

593

594 For *C. purpureus*, gametophyte DNA sequences including 18 samples at 6 locations are available
595 from Carey et al. (2021c). All RNA samples were taken from juvenile (protonema) and mature
596 (gametophore) tissues at the gametophyte stage. Gametophyte RNA sequences from 3 male-
597 female sibling pairs at 2 developmental stages (each with 3 biological replicates) are available
598 from Carey et al. (2021a). We generated sporophyte tissues for RNA sequencing using the same
599 methods and at the same time with Carey et al. (2021a). The sporophyte RNA sequences
600 included 26 samples from 6 crosses, the number of replicates per cross is between 2 and 10
601 (Table S3). We removed one sporophyte RNA sample from our analyses because it had low
602 coverage. The median read number for the sporophyte RNA sequences is 52,038,814 (range:
603 33,030,143 - 64,278,575) (Table S3).

604

605 *RNA Seq data and expression analysis*

606 For *R. hastatulus*, we used the genome assembly and genome annotation from Rifkin et al.
607 (2022) and assessed the quality of RNA reads using FastQC (Andrews 2010). In *C. purpureus*,
608 we used a hybrid genome consisting of the male R40 genome assembly (autosomes, the V sex
609 chromosome) and the U sex chromosome from the female GG1 genome assembly (Carey et al.
610 2021a). We trimmed adaptor sequences and filtered out low-quality reads using trimmomatic for
611 all RNA samples (Bolger et al. 2014), and trimmed polyG sequences using fastp for the
612 sporophyte RNA samples (Chen 2023). We used FastQC to assess the quality of RNA reads
613 before and after filtering (Andrews 2010). In *C. purpureus*, we excluded four sporophyte samples
614 for being an outlier in PCA from the majority of sporophyte samples, our final sample size for
615 sporophyte RNA samples is 21.

616

617 In both species, we mapped the RNA reads to the genome assembly using STAR in two pass
618 mode (Dobin et al. 2013, v2.7.6a). We sorted the alignment SAM files and added read groups
619 using PicardTools (http://broadinstitute.github.io/picard/). We generated raw read counts for each
620 gene using featureCount (Liao et al. 2014) and then normalized the raw read counts by library
621 depth using DESeq2 (Love et al. 2014). In the expression analysis of *C. purpureus*, we included
622 both male and female gametophyte samples for autosomal genes, and only included female or
623 male gametophyte samples for U and V genes, respectively. A gene was identified as expressed if
624 the mean normalized read count for either life stages was > 5; we only kept expressed genes in
625 the subsequent analyses. We calculated expression ratios between life stages for each gene by
626 dividing the mean normalized read counts at each life stage. A gene was considered tissue-
627 specific if: a) it was in the 10% or 90% quantiles of expression ratio among all significantly
628 differentially expressed genes (adjusted p-value < 0.1), and b) it had an absolute value of

629 log2FoldChange > 2. We performed GO enrichment for biased and specific genes using topGO,
630 with expressed genes from each species serving as the gene universe for comparison. (Alexa and
631 Rahnenfuhrer 2023). We divided the genes into four bins based on quantiles of log2FoldChange
632 for gametophyte-biased (log2FoldChange > 0) and sporophyte-biased (log2FoldChange > 0)
633 genes separately. We defined unbiased genes as the middle two bins with lowest absolute value
634 of log2FoldChange as well as an adjusted p >= 0.05. We divided the expressed genes into four or
635 ten quantiles of expression levels and used the same cutoff for later analyses.
636
637 *DNA Seq data and variant calling*
638 For *R. hastatulus*, we assessed the quality of DNA raw reads using FastQC (Andrews 2010) and
639 mapped the reads to the genome assembly using bwa-mem2 (Vasimuddin et al. 2019). We sorted
640 the BAM files using SAMtools (Danecek et al. 2021), added read groups, and removed PCR
641 duplicates using PicardTools (http://broadinstitute.github.io/picard/). We called SNPs jointly on
642 all samples using freebayes (Garrison and Marth 2012) with invariant sites saved in the output (-
643 -report-monomorphic). The variant and invariant sites were filtered separately using VCFtools
644 (Danecek et al. 2011), and concatenated together using BCFtools (Danecek et al. 2021). For both
645 variant and invariant sites we removed sites with a proportion of missing data > 0.8 or with a
646 mean read depth per site <10 or >40. For variant sites, we only kept bi-allelic SNPs with
647 genotype quality > 30.
648
649 For *C. purpureus*, we removed adaptor sequences using trimmomatic (LEADING:3
650 TRAILING:3 SLIDINGWINDOW:10:30 MINLEN:40) (Bolger et al. 2014), and performed
651 quality control using FastQC before and after trimming (Andrews 2010). We masked the V sex
652 chromosome when mapping female samples and masked the U sex chromosome when mapping

653 male samples using BEDTools maskfasta (Quinlan and Hall 2010). The read mapping and BAM
654 file processing methods are the same as described above in *R. hastatulus*. We called SNPs on all
655 samples jointly using BCFtools mpileup with a ploidy of 1 and the -B option (Danecek et al.
656 2021). We only kept the VCF for autosomal genes for the subsequent analysis, as the U and V
657 sex chromosomes have low recombination (Carey et al. 2021a, 2021c). We performed VCF
658 filtering using BCFtools (Danecek et al. 2021). We removed sites with a combined read depth <
659 5 and filtered variant and invariant sites separately. For variant sites, we kept bi-allelic SNPs and
660 removed sites with low quality or mapping score (QUAL<30 && MQ<30). Finally, we removed
661 sites with a proportion of missing data > 0.8 for both invariant and variant sites.
662
663 *Diversity statistics*
664 In both species we identified 0-fold and 4-fold degenerate sites in the genome using the script
665 codingSiteTypes.py(https://github.com/simonhmartin/genomics_general/blob/master/codingSite
666 Types.py, accessed in 2020). We intersected the list of 0-fold and 4-fold sites with the filtered
667 VCFs to generate 0-fold and 4-fold VCFs for population genetic analyses. The numbers of sites
668 (variant and invariant) in each VCF are 11,432,804 (0-fold) and 2,891,775 (4-fold) in *R.*
669 *hastatulus*; 17,394,329 (0-fold) and 4,759,819 (4-fold) in *C. purpureus*. We used pixy to
670 calculate the 0-fold and 4-fold average per site nucleotide diversity for each gene across the
671 genome (Korunes and Samuk 2021, v1.2.6.beta1). Only genes with a minimum of 50 sites for
672 both 4-fold and 0-fold nucleotide diversity were kept in the subsequent analyses.
673 We used the script popgenWindows.py
674 (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py, accessed
675 in 2020) to calculate Tajima's *D* for each gene based on the variant-only VCFs. We calculated
676 the weighted average nucleotide diversity for each bin based on the number of sites for each gene

677 (see supplementary methods). We bootstrapped the genes in each bin 1000 times to generate the
678 95% confidence intervals. We calculated the weighted average nucleotide diversity (mean and
679 95% CI) in the same way for gametophyte-specific, sporophyte-specific and unbiased genes. We
680 removed zeros for π_s and π_n to calculate the π_n/π_s ratio for each gene.
681
682 All DNA samples in *R. hastatulus* are female, this reduces the bias in nucleotide diversity
683 estimates due to the divergence between the X and Y sex chromosomes in males. In *C.*
684 *purpureus*, nucleotide diversity was estimated from 16 DNA samples (8 males and 8 females)
685 from 5 locations, see sample information from (2021a).
686
687 *Balancing selection scan*
688 We used BalLeRMix+ to perform genomic scans for balancing selection with B2 statistics and
689 default options (Cheng and DeGiorgio 2022). In *R. hastatulus*, the scans were performed on each
690 chromosome using the genetic positions. In *C. purpureus*, the scans were performed in 1 Mb
691 sliding windows based on the physical positions on each chromosome, using a default
692 recombination rate of 1 cM/Mb. The threshold for signature of balancing selection was set as the
693 5% quantile in the non-zero values of composite likelihood ratio. We calculated the maximum
694 CLR per gene. We identified the genes that had at least one site with strong signal of balancing
695 selection and also with evidence of expression in this study. We performed GO enrichment for
696 candidate balancing selection genes using topGO, with expressed genes as the gene universe for
697 comparison (Alexa and Rahnenfuhrer 2023). For candidate balancing selection genes with
698 gametophyte bias, we used gametophyte-biased genes as the comparison to test the functional
699 enrichment for balancing selection.
700

701 **Data availability**

702 Raw reads of *R. hastatulus* DNA and RNA Seq are available in NCBI SRA database under
703 BioProject PRJNA744278 (to be released on Aug 1, 2025 or upon publication). Raw reads of *C.*
704 *purpureus* sporophyte RNA Seq are available through JGI genome portal, see project IDs in
705 supplementary Table S3. Scripts used in this study are available at
706 https://github.com/imengyuan/lifestage_conflict (private repository, to be made public upon
707 review/publication).

708

709 **Acknowledgements**

710 We thank Bill Cole and Thomas Gludovacz for help with plant maintenance, Andrew Clemens
711 for assistance with *C. purpureus* sporophyte cultivation, and Tyler Kent for help with
712 interpolating recombination rate for *R. hastatulus*. This work was supported by NSERC
713 Discovery Grants awarded to JRS, SIW and SCHB, and NSF and NASA awards to SFM. The *C.*
714 *purpureus* sequencing (proposal: 10.46936/10.25585/60007219) conducted by the U.S.
715 Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of
716 Science User Facility, was supported by the Office of Science of the U.S. Department of Energy
717 operated under Contract No. DE-AC02-05CH11231.

References

Alexa A, Rahnenfuhrer J. 2023. topGO: Enrichment analysis for gene ontology. doi:10.18129/B9.bioc.topGO. https://bioconductor.org/packages/topGO/.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. Mol Biol Evol. 30(11):2475–2486. doi:10.1093/molbev/mst149.

Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano Kenji, Takano T, Yano Kentaro, Matsuoka M. 2015. De novo transcriptome assembly of a fern, *Lygodium japonicum,* and a web resource database, Ljtrans DB. Plant Cell Physiol. 56(1):e5–e5. doi:10.1093/pcp/pcu184.

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

Bitarello BD, Brandt DYC, Meyer D, Andrés AM. 2023. Inferring balancing selection from genome-scale data. Genome Biol Evol. 15(3):evad032. doi:10.1093/gbe/evad032.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective. J Exp Bot. 60(5):1465–1478. doi:10.1093/jxb/ern355.

Carey SB, Jenkins Jerry, Lovell John T., Maumus Florian, Sreedasyam Avinash, Payton Adam C., Shu Shengqiang, Tiley George P., Fernandez-Pozo Noe, Healey Adam, et al. 2021a. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. Sci Adv. 7(27):eabh2488. doi:10.1126/sciadv.abh2488.

Carey SB, Kollar LM, Mcdaniel SF. 2021b. Does degeneration or genetic conflict shape gene content on UV sex chromosomes? Bryophyte Divers Evol. 43(1):133–149. doi:10.11646/bde.43.1.11.

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

Cervantes S, Kesälahti R, Kumpula TA, Mattila TM, Helanterä H, Pyhäjärvi T. 2023. Strong purifying selection in haploid tissue–specific genes of scots pine supports the masking theory. Mol Biol Evol. 40(8):msad183. doi:10.1093/molbev/msad183.

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2(2):e107. doi:10.1002/imt2.107.

Cheng C, Kirkpatrick M. 2016. Sex-specific selection and sex-biased gene expression in humans and flies. Nachman MW, editor. PLOS Genet. 12(9):e1006170. doi:10.1371/journal.pgen.1006170.

Cheng X, DeGiorgio M. 2019. Detection of shared balancing selection in the absence of trans-species polymorphism. Mol Biol Evol. 36(1):177–199. doi:10.1093/molbev/msy202.

Cheng X, DeGiorgio M. 2020. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. Mol Biol Evol. 37(11):3267–3291. doi:10.1093/molbev/msaa134.

Cheng X, DeGiorgio M. 2022. BalLeRMix+: mixture model approaches for robust joint

764 identification of both positive selection and long-term balancing selection.
765 Bioinformatics. 38(3):861–863. doi:10.1093/bioinformatics/btab720.
766 Clark AG. 2002. Sperm competition and the maintenance of polymorphism. Heredity. 88(2):148–
767 153. doi:10.1038/sj.hdy.6800019.
768 Crow JF, Kimura M. 1965. Evolution in sexual and asexual populations. Am Nat. 99(909):439–
769 450. doi:10.1086/282389.
770 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
771 Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics.
772 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.
773 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
774 McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools.
775 GigaScience. 10(2):giab008. doi:10.1093/gigascience/giab008.
776 Dapper AL, Wade MJ. 2016. The evolution of sperm competition genes: the effect of mating
777 system on levels of genetic variation within and between species. Evol Int J Org Evol.
778 70(2):502–511. doi:10.1111/evo.12848.
779 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
780 TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15–21.
781 doi:10.1093/bioinformatics/bts635.
782 Dobzhansky T. 1970. Genetics of the evolutionary process. Columbia University Press.
783 Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, Andolfatto P, Coop G, Sella G. 2016.
784 A genomic map of the effects of linked selection in *Drosophila*. PLOS Genet.
785 12(8):e1006130. doi:10.1371/journal.pgen.1006130.
786 Field DL, Pickup M, Barrett SCH. 2012. The influence of pollination intensity on fertilization
787 success, progeny sex ratio, and fitness in a wind-pollinated, dioecious plant. Int J Plant
788 Sci. 173(2):184–191. doi:10.1086/663164.
789 Fishman L, Saunders A. 2008. Centromere-associated female meiotic drive entails male fitness
790 costs in monkeyflowers. Science. 322(5907):1559–1562. doi:10.1126/science.1161406.
791 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.
792 ArXiv12073907 Q-Bio. http://arxiv.org/abs/1207.3907.
793 Gerstein AC, Otto SP. 2009. Ploidy and the causes of genomic evolution. J Hered. 100(5):571–
794 581. doi:10.1093/jhered/esp057.
795 Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, Pantano L,
796 Bitarello BD, Izquierdo D, Noguera I, et al. 2019. Evolutionary and functional impact of
797 common polymorphic inversions in the human genome. Nat Commun. 10(1):4222.
798 doi:10.1038/s41467-019-12173-x.
799 Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014. Selection-driven evolution of
800 sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. Mol Biol
801 Evol. 31(3):574–583. doi:10.1093/molbev/mst226.
802 Gutiérrez-Valencia J, Fracassetti M, Horvath R, Laenen B, Désamore A, Drouzas AD, Friberg
803 M, Kolář F, Slotte T. 2022. Genomic signatures of sexual selection on pollen-expressed
804 genes in *Arabis alpina*. de Meaux J, editor. Mol Biol Evol. 39(1):msab349.
805 doi:10.1093/molbev/msab349.
806 Haig D, Wilczek A. 2006. Sexual conflict and the alternation of haploid and diploid generations.
807 Philos Trans R Soc B Biol Sci. 361(1466):335–343. doi:10.1098/rstb.2005.1794.
808 Haldane JBS. 1932. The causes of evolution. Princeton (NJ): Princeton University Press.
809 Harrison MC, Mallon EB, Twell D, Hammond RL. 2019. Deleterious mutation accumulation in
810 *Arabidopsis thaliana* pollen genes: a role for a recent relaxation of selection. Genome

811 Biol Evol. 11(7):1939–1951. doi:10.1093/gbe/evz127.
812 Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant
813 Physiol. 132(2):640–652. doi:10.1104/pp.103.020925.
814 Hormaza JI, Herrero M. 1992. Pollen selection. Theoret Appl Genetics. 83(6–7):663–672.
815 Huang K, Rieseberg LH. 2020. Frequency, origins, and evolutionary role of chromosomal
816 inversions in plants. Front Plant Sci. 11:296. doi:10.3389/fpls.2020.00296.
817 Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. 2014. Exposing synonymous
818 mutations. Trends Genet. 30(7):308–321. doi:10.1016/j.tig.2014.04.006.
819 Immler S, Arnqvist G, Otto SP. 2012. Ploidally antagonistic selection maintains stable genetic
820 polymorphism: ploidally antagonistic selection. Evolution. 66(1):55–65.
821 doi:10.1111/j.1558-5646.2011.01399.x.
822 Johnson MG, Shaw AJ. 2016. The effects of quantitative fecundity in the haploid stage on
823 reproductive success and diploid fitness in the aquatic peat moss *Sphagnum*
824 *macrophyllum*. Heredity. 116(6):523–530. doi:10.1038/hdy.2016.13.
825 Kasimatis KR, Nelson TC, Phillips PC. 2017. Genomic signatures of sexual conflict. J Hered.
826 108(7):780–790. doi:10.1093/jhered/esx080.
827 Kidwell JF, Cleggi MT, Stewart FM. 1976. Regions of stable equilibria for models of differential
828 selection in the two sexes under random mating. Genetics. 85(1):171–183.
829 Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation.
830 Genetics. 173(1):419–434. doi:10.1534/genetics.105.047985.
831 Klodová B, Fíla J. 2021. A decade of pollen phosphoproteomics. Int J Mol Sci. 22(22):12212.
832 doi:10.3390/ijms222212212.
833 Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? Nature. 351(6324):314–
834 315. doi:10.1038/351314a0.
835 Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence
836 in the presence of missing data. Mol Ecol Resour. 21(4):1359–1368. doi:10.1111/1755-
837 0998.13326.
838 Kruglyak L, Beyer A, Bloom JS, Grossbach J, Lieberman TD, Mancuso CP, Rich MS, Sherlock
839 G, Kaplan CD. 2023. Insufficient evidence for non-neutrality of synonymous mutations.
840 Nature. 616(7957):E8–E9. doi:10.1038/s41586-023-05865-4.
841 Lankinen A, Kiboi S. 2007. Pollen donor identity affects timing of stigma receptivity in *Collinsia*
842 *heterophylla* (Plantaginaceae): a sexual conflict during pollen competition? Am Nat.
843 170(6):854–863. doi:10.1086/522839.
844 Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at
845 synonymous sites in *D. melanogaster*. PLOS Genet. 9(5):e1003527.
846 doi:10.1371/journal.pgen.1003527.
847 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for
848 assigning sequence reads to genomic features. Bioinforma Oxf Engl. 30(7):923–930.
849 doi:10.1093/bioinformatics/btt656.
850 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
851 RNA-seq data with DESeq2. Genome Biol. 15(12):550. doi:10.1186/s13059-014-0550-8.
852 Lu Y. 2011. RNA isolation from *Arabidopsis* pollen grains. Bio-Protoc. 1(9):e67.
853 doi:10.21769/BioProtoc.67.
854 Mank JE. 2017. Population genetics of sexual conflict in the genomic era. Nat Rev Genet.
855 18(12):721–730. doi:10.1038/nrg.2017.83.
856 McDaniel SF. 2023. Divergent outcomes of genetic conflict on the UV sex chromosomes of
857 *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.
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.
Slotte T, Bataillon T, Hansen TT, St. Onge K, Wright SI, Schierup MH. 2011. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. Genome Biol Evol. 3:1210–1219. doi:10.1093/gbe/evr094.
Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ. 2011. Generation-biased gene expression in a bryophyte model system. Mol Biol Evol. 28(1):803–812. doi:10.1093/molbev/msq254.

905 Szövényi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. 2013. Selection is no more
906 efficient in haploid than in diploid life stages of an angiosperm and a moss. Mol Biol
907 Evol. 30(8):1929–1939. doi:10.1093/molbev/mst095.
908 Travers SE, Mazer SJ. 2001. Trade-offs between male and female reproduction associated with
909 allozyme variation in phosphoglucoisomerase in an annual plant (*Clarkia unguiculata*:
910 Onagraceae). Evolution. 55(12):2421–2428. doi:10.1111/j.0014-3820.2001.tb00757.x.
911 Urrutia AO, Hurst LD. 2003. The signature of selection mediated by expression on human genes.
912 Genome Res. 13(10):2260–2264. doi:10.1101/gr.641103.
913 Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-
914 MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed
915 Processing Symposium (IPDPS). p. 314–324.
916 Walbot V, Evans MMS. 2003. Unique features of the plant life cycle and their consequences. Nat
917 Rev Genet. 4(5):369–379. doi:10.1038/nrg1064.
918 Wright AE, Fumagalli M, Cooney CR, Bloch NI, Vieira FG, Buechel SD, Kolm N, Mank JE.
919 2018. Male-biased gene expression resolves sexual conflict through the evolution of sex-
920 specific genetic architecture. Evol Lett. 2(2):52–61.
921 Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. Nat Rev Genet.
922 16(7):409–420. doi:10.1038/nrg3950.
923 Zhou Z, Dang Y, Zhou M, Li L, Yu C, Fu J, Chen S, Liu Y. 2016. Codon usage is an important
924 determinant of gene expression levels largely through its effects on transcription. Proc
925 Natl Acad Sci. 113(41):E6117–E6125. doi:10.1073/pnas.1606724113.