Reproductive proteins evolve faster than non-reproductive proteins among *Solanum* species

Leonie C. Moyle[1*], Meng Wu[1], Matthew J. S. Gibson[1]

[1] Dept of Biology, Indiana University, Bloomington, Indiana, USA

*** Correspondence:**
lmoyle@indiana.edu

Keywords: female, male, molecular evolution, reproductive protein, sexual selection, *Solanum*, tomato

Abstract

Elevated rates of evolution in reproductive proteins are commonly observed in animal species, and are thought to be driven by the action of sexual selection and sexual conflict acting specifically on reproductive traits. Whether similar patterns are broadly observed in other biological groups is equivocal. Here we examine patterns of protein divergence among wild tomato species (Solanum section Lycopersicon), to understand forces shaping the evolution of reproductive genes in this diverse, rapidly evolving plant clade. By comparing rates of molecular evolution among loci expressed in reproductive and non-reproductive tissues, our aims were to test if: a) reproductive-specific loci evolve more rapidly, on average, than non-reproductive loci; b) 'male'-specific loci evolve at different rates than 'female'-specific loci; c) genes expressed exclusively in gametophytic (haploid) tissue evolve differently from genes expressed in sporophytic (diploid) tissue or in both tissue types; and d) mating system variation (a potential proxy for the expected strength of sexual selection and/or sexual conflict) affects patterns of protein evolution. We observed elevated evolutionary rates in reproductive proteins; however this pattern was most evident for female- rather than male-specific loci, both broadly and for individual loci inferred to be positively selected. These elevated rates might be facilitated by greater tissue-specificity of reproductive proteins, as faster rates were associated with more narrow expression domains. In contrast we found little evidence that evolutionary rates are consistently different in loci experiencing haploid selection (gametophytic-exclusive loci), or in lineages with quantitatively different mating systems. Overall while reproductive protein evolution is generally elevated in this diverse plant group, specific patterns of protein evolution are more complex than those reported in other (largely animal) systems, and include a more prominent role for female-specific loci among adaptively evolving genes.

1 Introduction

The rapid evolution of reproductive proteins has been observed across many different animal species, including in groups as diverse as marine invertebrates, insects, and mammals (Swanson and Vacquier 2002; Clark et al. 2006; Turner and Hoekstra 2008). This pattern is especially well established for proteins involved in male-specific functions, such as seminal fluid proteins, and sperm-egg and sperm-reproductive tract (RT) interactions. Although many factors can influence molecular evolutionary rates, two specific evolutionary forces—sexual selection and sexual conflict—have been proposed as primary drivers of this accelerated evolution, because both processes are expected to differentially affect reproductive traits and the proteins that underlie them. That is, traits mediating male-male competition and male-female mate choice experience unique selection to maximize mating and reproductive opportunities (Andersson 1994), and often appear to evolve rapidly between species (Ritchie 2007); therefore, the loci underpinning these reproductive traits might also be expected to display accelerated evolution and protein divergence between species (Swanson and Vacquier 2002, Clark et al. 2006, Vacquier and Swanson 2011). This inference is supported by observations of especially exaggerated protein evolution in male-exclusive and male-biased loci (collated in Dapper and Wade 2020), the sex that is usually subject to more intense sexual selection (Andersson 1994). Importantly, if sexual selection or conflict are the most critical factors driving rapid reproductive protein evolution, this pattern should be observable in other groups of organisms that also experience these selective forces. Nonetheless, whether reproductive proteins routinely evolve rapidly in non-animal species remains unclear.

Among these other groups, flowering plants (angiosperms) have numerous reproductive traits that can influence the operation and intensity of intrasexual competition and/or mate choice (Lloyd and Webb 1977, Willson 1979, Stephenson and Bertin 1983, Delph and Ashman 2006, Moore and Pannell 2011)—the two components of sexual selection. These include pollinator attraction traits, pollen competition traits, female reproductive tract ("pistil") traits (that influence pollen performance and fertilization success), and seed maturation traits (that can be used to exercise mate choice via selective abortion) (Willson 1994, Delph and Haven 1998, Skogsmyr and Lankinen 2002). Accordingly, angiosperms may also be expected to exhibit elevated evolutionary rates in loci underpinning these traits, similar to those inferred in animals (Clark et al. 2006). Indeed, of the few studies in angiosperms, some analyses have found evidence for faster protein evolution in reproductive loci (Szovenyi et al. 2013, Harrison et al. 2019), or greater representation of rapidly evolving genes among loci expressed in certain reproductive tissues (Gossman et al. 2016). However, others have shown more equivocal patterns (e.g. Gossman et al. 2014)—for example, sex-biased genes do not appear to evolve faster than non-sex-biased genes in several dioecious plant species (reviewed in Muyle 2019)—or even inferred greater levels of constraint on reproductive-specific proteins (e.g., Darolti et al. 2018).

These more complex findings could reflect the influence of additional factors on rates of protein evolution in plants, including the opportunity for haploid selection and the complex effects of mating system variation. In the first case, plants often express a substantial

proportion of their genome during the haploid phase of the life cycle (Hafidh et al. 2016), thereby exposing these 'gametophytic' loci to haploid selection. In flowering plants, gametophytic traits are largely reproductive and include 'male' functions such as pollen tube germination, growth, and interactions with the female pistil, and 'female' functions such as ovule signaling (Hafidh et al. 2016, Mizuta and Higashiyama 2018). Compared to diploid-expressed loci, traits that rely on haploid gene function are expected to experience both stronger purifying selection against deleterious alleles, and elevated fixation rates (stronger positive selection) for advantageous alleles, because both deleterious and advantageous alleles will be visible to haploid selection regardless of dominance (Charlesworth and Charlesworth 1992, Walsh and Charlesworth 1992, Immler and Otto 2018, Immler 2019). Moreover, gametophytic-exclusive genes could also evolve differently from loci that are expressed in both gametophytic and sporophytic (diploid) tissue, if the latter also experience additional constraints, or even antagonistic effects, because of their expression during both phases of the life cycle (Immler and Otto 2018).

In addition to haploid selection, plants also often exhibit substantial variation in mating system, including among closely related species. This variation is predicted to have diverse effects on the nature, timing, and strength of selection, acting both on reproductive genes specifically, and across the genome more generally. In the specific case of reproductive genes, because the mating system affects the number of reproductive partners an individual experiences, it can directly influence the intensity of intrasexual competition, the opportunity for mate choice, and the magnitude of conflict over reproductive decisions (Delph and Havens 1998; Skogsmyr and Lankinen 2002; Brandvain and Haig 2005, Clark et al. 2006, Mazer et al. 2010). Selection on traits affecting sexual competition and/or sexual conflict will be greatest in obligately outcrossing individuals, but weak or absent in obligately self-fertilizing species where the reproductive interests of both parents align perfectly (because they are the same individual). Even intermediate rates of self-fertilization can diminish the strength of sexual selection and thereby reduce, for example, the advantage of pollen competition traits (e.g. Mazer et al. 2018) and the spread of superior pollen-expressed genes (e.g. Peters and Weis 2018). Mating system—specifically variation in the frequency of selfing—is also proposed to have more global effects on the efficacy of selection across the genome, regardless of whether genes have reproductive functions. Because selfing reduces the effective population size (*Ne*), the overall efficacy of selection—both against deleterious alleles, and in favor of advantageous alleles—is expected to be reduced in selfing compared to outcrossing lineages (Charlesworth et al. 1993, Charlesworth and Wright 2001, Mattila et al. 2018). Conversely, because selfing increases homozygosity, it can also reduce the genetic load in a population—by increasing the exposure of strongly deleterious recessive alleles to selection—as well as increase the early efficacy of selection on recessive advantageous alleles by exposing them more rapidly to positive selection (Mattila et al. 2018). Together with specific effects on reproductive genes, these complex and sometimes antagonistic genome-wide effects might contribute to the heterogeneous inferences that have emerged from studies of reproductive protein evolution in plants to date.

In this study, we examined genome-wide patterns and rates of protein evolution among four closely-related species from the diverse plant genus Solanum. Our aims were to compare

patterns of molecular evolution in loci expressed in different classes of tissue in order to evaluate whether: a) reproductive-specific proteins evolve more rapidly, on average, than non-reproductive proteins; b) 'male'-specific loci evolve at different rates than 'female'-specific loci; c) genes expressed exclusively in gametophytic tissues evolve differently from genes expressed in sporophytic (diploid) tissue or in both tissue types; and d) mating system variation (a potential proxy for the expected strength of sexual selection and/or sexual conflict) affects patterns of protein evolution. In addition, to help assess underlying causes of detected patterns, we also evaluated the influence of the magnitude and breadth of gene expression on observed rates of protein evolution. Our goals were to understand the forces shaping the molecular evolution of reproductive proteins in this plant group, and whether this differs from evolutionary patterns in other protein types and/or those proposed to shape reproductive protein evolution in animal systems.

2 Materials and Methods

2.1 Study Species

Our analyses used gene expression and sequence data from each of four species within the wild tomato group (Solanum section Lycopersicum): *S. lycopersicum* (domesticated tomato; LA3475) and three wild relatives—*S. pimpinellifolium* (LA1589), *S. pennellii* (LA0716), and *S. habrochaites* (LA1777). (LA#### refers to the specific germplasm accession ID, as described in tgrc.ucdavis.edu, from which all or most of the data were obtained; Table S1.) This group is a clade of 12 closely-related diploid wild species along with the domesticated tomato, within the hyperdiverse (~1300 species) plant genus *Solanum* (Sarkinen et al. 2013). All 12 wild tomato species arose within the last ~2.5 million years, consistent with recent rapid speciation in this clade (Pease et al. 2016a).

Mating system varies substantially across species in *Solanum* (Goldberg et al. 2010) and our four focal species differ in their mating system, most notably in whether species are historically self-incompatible (SI: *S. pennellii* and *S. habrochaites*) versus self-compatible (SC: *S. lycopersicum* and *S. pimpinellifolium*). These species also have large differences in their genetic effective population sizes (Ne)—for example, as estimated from average nucleotide polymorphism (Pease et al. 2016a)—that are consistent with these SI vs SC mating system differences. In *Solanum*, genetic self-incompatibility (the inferred ancestral state) is based on molecular interactions between proteins expressed in growing pollen tubes and pistil-expressed SRNase and other known proteins (Cruz-Garcia et al. 2003; Takayama and Isogai 2005). SI is observed as arrest of pollen tube growth in the female reproductive tract (the pistil). Self-compatible lineages have lost function in one or more SI-mediating proteins (Stone 2002), thereby allowing individuals to be fertilized by their own pollen, in addition to outcrossing. As with other angiosperms (Sicard and Lenhard 2011), transitions to SC within *Solanum* are often followed by other morphological changes in floral size and structure, that further increase the frequency of self-pollination and reduce the number/diversity of mating partners (Rick 1979). In our four focal species, SI species have larger flowers, longer reproductive tracts, and more

178 pollen production on average, in comparison to SC species (Vosters et al. 2014), even though
179 the transition to SC likely occurred within the last ~0.5MY (the estimated age of the clade
180 containing the two SC species here; Pease et al. 2016a). Both differences in the capacity to
181 reject self pollen (i.e. SI versus SC), and these morphological differences, could have large
182 effects on the average diversity of mating partners, and therefore the opportunity for and
183 magnitude of intraspecific sexual selection and sexual conflict (Willson 1994, Delph and Havens
184 1998, Clark et al. 2006).
185
186 **2.2 Quantifying gene expression and determining tissue-specificity**
187
188 Our analyses used gene expression (RNAseq) data collected individually from leaf (up to five
189 developmental stages), root, seed, vegetative meristem, stem, flower (several stages), style,
190 pollen, and ovule tissues in each accession. RNAseq data were obtained from seven publically
191 available SRA projects (Table S1), mostly drawn from three previously published analyses
192 (Koenig et al. 2013, Ichihashi et al. 2014, Pease et al. 2016b), except data from ovule tissue
193 which has not been previously published. Note that style and pollen data were not available
194 specifically for *S. pimpinellifolium*. Sources of each dataset are described in detail in Table S1.
195 Procedures for generating the ovule RNAseq data are described in the supplementary text. Each
196 tissue had 1-6 replicates (generally 3) per accession (Table S1).
197
198 For each library within each tissue, we processed the raw reads by filtering adapter sequences
199 or low quality bases with TRIMMOMATIC (Bolger et al. 2014). Reads were then mapped against
200 tomato reference genome (ITAG 2.4) using STAR (Dobin et al. 2013). Numbers of reads mapped
201 onto genic regions were estimated with FEATURECOUNTS (Liao et al. 2014). We normalized the
202 read counts from each library by TPM (transcripts per million) and then calculated the mean
203 normalized read counts for each tissue of each species.
204
205 We first classified loci into three general classes of gene—reproductive (RP), vegetative (VG),
206 and general (GR) (i.e., expressed in both reproductive and non-reproductive/vegetative tissue
207 types). Reproductive genes were required to be expressed (TPM>2) in at least one
208 reproductive tissue (i.e. style, pollen, ovule), and to have no or trace expression (TPM <0.5) in
209 any of the remaining tissues (i.e. leaf, root, stem, seed, root, vegetative meristem), in at least
210 two of the three investigated species for which we had all tissues available (i.e. *S. lycopersicum*,
211 *S. pennellii*, and *S. habrochaites*). (Because *S. pimpinellifolium* lacked some of the tissue types in
212 our RNA-seq data, we did not include this species in the filtering/determination of tissue-
213 specificity.) Vegetative genes were defined similarly by requiring expression in at least one
214 vegetative (non-reproductive) tissue, and no/trace expression in any reproductive tissue.
215 Generally-expressed (GR) genes were expressed in at least one reproductive tissue and also in
216 at least one vegetative tissue.
217
218 Using similar criteria, we also identified a set of tissue-specific genes, defined as those only
219 expressed in one specific tissue (TPM>2) and not in any other tissue (TPM<0.5) in at least two
220 out of the three species for which we had all tissue types (i.e. *S. lycopersicum*, *S. pennellii*, and
221 *S. habrochaites*). For these analyses, we focused on specific tissues (i.e. leaf, root, stem, seed,

222 vegetative meristem, style, pollen, ovule) for which we had sufficient sampling for at least three
223 species (Table S1). Finally, using the same criteria and the expression domains in the three
224 species for which we had all tissues represented, we also classified transcripts as gametophytic-
225 exclusive (pollen and/or ovule expression only) versus sporophytic-exclusive versus both
226 gametophytic and sporophytic (pollen and/or ovule expression, and at least one other diploid
227 tissue), as well as ovule-specific versus expressed in both ovules and at least one sporophytic
228 (diploid) tissue, and pollen-specific versus expressed in both pollen and at least one sporophytic
229 (diploid) tissue.
230
231 **2.3 Estimating rates of protein evolution**
232
233 As with prior studies (Clark et al. 2006 and references therein), we used dN/dS—the ratio of the
234 number of non-synonymous substitutions per non-synonymous site to the number of
235 synonymous substitutions per synonymous site—as our estimate of protein evolution. Although
236 dN/dS is usually <1—because most functional loci experience some constraint on non-
237 synonymous changes—larger values of dN/dS are consistent with faster protein evolution. We
238 estimated dN/dS for each locus in our dataset, to calculate mean rates of protein evolution in
239 different classes of genes, and to identify the set of loci with estimated dN/dS>1—a pattern
240 consistent with positive selection for protein evolution at that locus. In addition, to compare
241 the overall dN/dS for loci on SI versus SC branches, we compared estimates from sequences
242 that were concatenated across groups of genes (described below).
243
244 To generate the sequence alignments for these analyses, we combined the RNA-seq data used
245 for quantifying gene expression (above) with whole-transcriptome data from the same four
246 accessions, previously generated in Pease et al. (2016a) or publically available (Table S2); the
247 latter dataset allowed us to augment our RNAseq data with additional sequences for loci whose
248 expression was not detected for one or more of our four species. These reads were aligned
249 against the tomato reference genome by STAR. The aligned data were processed using
250 SAMTOOLS followed by MVFTOOLS to generate the sequence alignments of orthologous genes. We
251 only retained transcript sequence alignments longer than 200 bps after removing gaps and
252 indels, leaving us with 21,216 sequence alignments (loci) for downstream analyses.
253
254 For each locus, we used PAML (Yang 2007) to estimate dN/dS across the branches of a given
255 phylogenetic topology among the four investigated species, using the input species tree
256 ((lycopersicum, pimpinellifolium), (pennellii, habrochaites)). To do so we used the one-ratio
257 (M0) codeml model in PAML, that generates a single dN/dS estimate across all branches for each
258 locus (that is, dN/dS is fixed at one ratio for each gene). Outputs from this model were used to
259 compare the mean and distribution of dN/dS for genes among three broad categories
260 (reproductive (RP), vegetative (VE), and general (GR)), as well as between each of six tissue-
261 specific groups classified by their expression in only a single tissue (pollen, style, ovule, leaf,
262 root, or stem; vegetative meristem and seed were excluded from the latter analyses because
263 they had no exclusively tissue-specific genes), and between loci that had gametophytic versus
264 sporophytic, or gametophytic+sporophytic, domains of gene expression. We also compared loci

265 expressed in ovules only versus expressed in ovules and at least one sporophytic (diploid)
266 tissue, or in pollen only versus expressed in pollen and at least one sporophytic tissue.
267
268 Second, in addition to locus-by-locus dN/dS comparisons, we also evaluated overall dN/dS
269 estimates along SC branches (those leading to *S. lycopersicum* and *S. pimpinellifolium*) in
270 comparison to those along SI branches (those leading to *S. pennellii* and *S. habrochaites*).
271 Because the four species in our analysis are very closely related, the branch lengths for each
272 class of lineage (especially for SC lineages) were too short to reliably estimate lineage-specific
273 dN/dS for each locus individually. Therefore, for these analyses we concatenated all loci within
274 each class of genes to be compared, and then estimated dN/dS across the concatenated set of
275 loci for SI and SC branches separately. For example, for reproductive (RP) loci, we concatenated
276 all RP loci and estimated dN/dS across the entire concatenated set, allowing for different dN/dS
277 estimates (variable rates) on SI versus SC branches. Because of the small numbers of loci
278 concatenated for some of our individual tissues, we limited SI vs SC branch-specific estimates to
279 our three broad classes of loci (RP, VG, and GR) only.
280
281 Because the phylogeny of our species contains one internal branch, for these analyses we used
282 two alternative multi-rate (M2) codeml models in PAML. For the first, three-ratio, model we
283 allowed the internal branch to have a rate that differed from the tip branches leading either to
284 SC species, or to SI species. For the second, two-ratio, model, we classified the internal branch
285 as 'SI' (i.e. dN/dS was estimated separately for tip branches leading to SC species, versus a
286 background rate for the tip branches leading to SI species plus the internal branch). This
287 approach was used to estimate branch-specific rates (dN/dS estimates) for SI versus SC
288 branches for each of our three general classes of loci (RP, VG, and GR). Note we report three-
289 ratio results in the main text, and two-ratio results in the supplement.
290
291 **2.4 Comparing protein evolution rates**
292
293 All statistical analyses were performed in R version 3.6.3 (R Core Team 2015). For all
294 comparisons, we first removed genes showing unusually high values (dN/dS >10) from the PAML
295 output, as these were due to estimation limitations within PAML (e.g. a lack of synonymous
296 substitutions across the branch(es) used for estimation) or otherwise likely resulted from poor
297 sequence alignment.
298
299 To compare the mean rate of per locus protein evolution (dN/dS) between different classes of
300 loci, we fit generalized linear models (GLMs). Shapiro-Wilk tests of normality and quantile-
301 quantile plots were used to assess normality in the distribution of dN/dS; in all categories,
302 dN/dS was heavily skewed towards lower values and zero-inflated (reflecting the fact that most
303 expressed genes are under purifying selection). To accommodate this in our models, we assume
304 a gamma residual distribution with an identity link function, rather than the standard Gaussian
305 assumption, after confirming the suitability of this distribution with maximum likelihood (using
306 the fitdistrplus R package). Following GLMs, to compare the mean estimated dN/dS between
307 specific classes of loci, we made pairwise comparisons using Tukey post-hoc tests, adjusting p-
308 values for multiple comparisons. In addition, we compared the proportion of positively selected

genes (i.e. dN/dS>1) between different classes of loci by performing Chi-Squared tests of independence and, for specific tests of enrichment in single categories (e.g., in RP relative to GR), one-sided Fisher's exact tests.

To assess evidence for differences in dN/dS between mating systems, we aimed to compare dN/dS estimates from concatenated data, as calculated separately for SI and SC branches. Both biological and technical factors make these comparisons challenging to interpret. Biologically, overall differences in Ne and therefore in the expected efficacy of positive and purifying selection, could generate global differences in rates of protein evolution between SC and SI lineages (see Introduction), that are unrelated to their potential differences in the historical strength of sexual selection. Technically, in our specific dataset, the two SC species are much more closely related than the two SI species, so there are large differences in branch lengths (number of substitutions) over which we are estimating dN/dS values for SI versus SC branches. Therefore, to assess whether exclusively reproductive (RP) loci had a different pattern of molecular evolution on SI versus SC branches in comparison to other loci, we compared the observed difference in dN/dS between SI and SC branches for RP genes to an estimate of the 'baseline' genome-wide difference in dN/dS between SI and SC branches. This baseline was described by the SI vs SC difference in our set of generally expressed genes (GR), which is expected to be due to factors unrelated to variation in sexual selection. Because we have only a single estimate (from the concatenated data) for each class of gene for each branch type, standard parametric tests can't be used. Instead we used bootstrap resampling to estimate the variance around the mean SI vs SC difference in dN/dS in the GR dataset, and then to evaluate whether the observed SI vs SC difference in RP genes falls outside the range of this variance (i.e. is larger or smaller than the 95% CI of this distribution). For each bootstrap replicate (N=1000 replicates), we re-sampled 500 loci at random from the GR dataset, estimated dN/dS for SI and SC branches and the difference between these two estimates. Our observed dN/dS difference (SC-SI rate) was compared to the resulting distribution of 1000 replicate dN/dS differences from GR loci.

2.5 Assessing the effect of gene expression level on molecular evolutionary rates

Numerous studies have shown that rates of protein evolution can be influenced by the magnitude and breadth of gene expression, such that highly expressed genes and/or genes expressed in a broader number of tissues tend to have systematically lower estimated dN/dS (Meisel 2011). To assess the effect of the magnitude and breadth of gene expression on our inferences, we also compared average dN/dS values between genes expressed in different numbers of tissues (ranging from 1 (tissue-specific loci) through to 8 (expressed in all tissues examined)), using the criteria that a gene be expressed >2 TPM in a specific tissue in all species for which we had data on that tissue-type. Within each of the broad classes of RP, VG, and GR loci, we also assessed the quantitative relationship between mean gene expression level (averaged across all tissues in which a locus was expressed) and estimated dN/dS. For this analysis we constructed a multiple regression model of dN/dS (square root transformed to approximate normality), including both mean gene expression (TPM), tissue class (RP, VG, or GR), and their interaction term as independent variables.

2.6 Assessing enrichment of functional categories among adaptively evolving loci

To test for functional enrichment among our loci with dN/dS > 1, we performed a Gene Ontology (GO) term enrichment analysis using Panther version 14 (Mi et al., 2019; Ashburner et al., 2000; The Gene Ontology Consortium, 2019). We performed separate tests for GR, VE, and RP tissue categories, as these classes had large sample sizes appropriate for tests of enrichment. Loci were evaluated using both the biological process and molecular function annotation data sets, and significance was determined using Fisher's Exact tests with false discovery rate correction.

3 Results

3.1 Reproductive genes evolve at modestly faster rates, and have proportionally more genes under positive selection

Reproductive (RP) genes had an elevated mean rate of protein evolution (0.307±0.014) compared to loci expressed in both reproductive and non-reproductive (general/GR) tissues (0.271±0.003) (Posthoc Z=2.548, p = 0.0108) (Figure 1; Table 1; Table S3, Table S4). The mean dN/dS for vegetative-specific (VG) loci (0.278±0.011) fell between these two other classes of genes and did not differ from either of them (Table 1). In addition, the proportion of genes with dN/dS>1 differed by group ($X^2(2)$= 7.9696, p=0.0186). A significantly greater proportion of reproductive loci (28 of 670, or ~4.17%) are inferred to be evolving adaptively compared to general loci (2.45%) (Fisher's exact test, p=0.0067; Table 1, Table S4). The proportion of vegetative loci evolving adaptively (2.81%) was marginally lower than reproductive loci (Fisher's exact test, p=0.093), and did not differ from general loci (Table 1; Table 4).

3.2 Female- rather than male-specific genes evolve the most rapidly

Comparisons among tissue-specific genes indicated no evidence that pollen-specific loci evolve more rapidly than loci from female-derived tissues (style and ovule) (Figure 2; Table 2: Table S5). Instead, style- and ovule-specific loci had the highest mean dN/dS values of all tissue groups (0.67±0.17 and 0.34±0.08 respectively), and posthoc pairwise tests indicated that style mean rate was significantly greater than rates in root-specific and leaf-specific loci, and ovule mean rate was significantly greater than leaf mean rate (Figure 2; Table 2; Table S6). In comparison, pollen loci had a mean rate (0.25±0.10) that was intermediate between these groups of tissue-specific genes, and statistically indistinguishable from each other tissue (Table S6). Note that many fewer loci were available for tissue-specific comparisons, especially leaf- and stem-exclusive loci (Table 2), and two tissues—vegetative meristem and seed—had no detected tissue-exclusive loci.

Similar to comparisons of mean rates, the proportion of genes with dN/dS>1 differed by tissue ($X^2(5)$=33.707, p=2.723 x 10^{-6}). Moreover, this proportion was significantly higher in style

397 tissue—where 6 of 25 loci were inferred to be evolving adaptively—compared to all other
398 tissues (Fisher's exact test, p= 5.046 x 10^{-5}; Table 2). For ovule-exclusive loci, the proportion
399 adaptive evolving loci (3 of 71) was also greater than in all other tissues (excluding style-specific
400 loci) (Fisher's exact test, p= 0.0376) (Table 2).
401
402 **3.3 Gametophytic-exclusive proteins do not evolve more rapidly than sporophytic-**
403 **exclusive proteins, or proteins expressed in both tissue types**
404
405 Although the mean rate of protein evolution in gametophytic genes (0.33±0.04) was slightly
406 higher than the rate of protein evolution in sporophytic-exclusive loci (0.30±0.06), this
407 difference was not significant (t=-0.452, P=0.588) (Table 3; Table S7). Gametophytic-exclusive
408 proteins also do not evolve faster than those expressed in at least one gametophytic and one
409 sporophytic tissue (0.27±0.03) (t=-1.498, P=0.134). Similar to mean rate comparisons, the
410 proportion of gametophytic loci with dN/dS>1 (3 of 86, or 3.49%) did not differ from this
411 proportion in sporophytic-exclusive genes (68 of 2033, or 3.34%) (Fisher's exact test, p=0.557),
412 and was slightly but not statistically larger than this proportion in gametophytic+sporophytic
413 genes (383 of 15778, or 2.43%) (Fisher's exact test, p=0.283).
414
415 As observed in the general gametophytic comparisons, mean rate of protein evolution in ovule-
416 exclusive loci (0.34±0.46) was slightly higher than the rate of protein evolution in loci expressed
417 in ovules and at least one sporophytic tissue (0.29±0.32), but this difference was not significant
418 (GLM t=1.212, p=0.225; Table S5). The proportion of ovule-exclusive loci with dN/dS>1 (3 of 71,
419 or 4.23%) also did not differ from the proportion of ovule+sporophytic loci with dN/dS>1 (225
420 of 7884, or 2.85%) (Fisher's exact test, p=0.854). Neither the mean rate of protein evolution
421 (GLM t=-0.037, p=0.971; Table S7) nor the proportion of loci with dN/dS>1 (Fisher's exact test,
422 p=0.832) differed between pollen-exclusive loci and loci expressed in pollen and at least one
423 sporophytic tissue (Table 3).
424
425 **3.4 Patterns of protein evolution show mating-system effects that are not specific to**
426 **reproductive loci**
427
428 Consistent with genome-wide effects of lower Ne in our self-compatible lineages, rates of
429 protein evolution were inferred to be higher overall on self-compatible branches in both GR
430 and RP classes of loci, although not for VG loci (Table 4); in the latter case, VG rates were
431 numerically higher on SI branches. Results were similar for two-ratio models (Table S8). The
432 observed difference in rates between SC and SI branch for RP loci (0.0338) did not fall outside
433 the 95% CI of bootstrapped values (0.0007-0.0797) from the broader class of GR genes; instead
434 33% of bootstrapped samples had a dN/dS difference smaller than observed for RP loci.
435
436 **3.5 Higher rates of protein evolution are consistently associated with greater tissue-**
437 **specificity**
438
439 Across loci classified according to the number of tissues in which they were expressed at >2
440 TPM (from 1 to 8 tissues), mean dN/dS was highest in loci expressed in a single tissue and

consistently declined across loci expressed in increasingly more tissues (Figure 3; Table 5; Table S9, S10). Similarly, both mean expression level and tissue class (GR, RP, VG) affected the per locus estimates of protein evolution (dN/dS), as did their interaction (Table 6), indicating highly expressed loci generally have lower average rates of protein evolution but also that this relationship varies depending upon the broad tissue class. The relationship between dN/dS and mean expression level was negative for generally-expressed (GR) loci, however the 95% CI for this slope overlapped zero for both RP and VG loci, indicating weaker relationships for these less broadly expressed loci (Figure S1; Table S11). Classes of genes also differed in their overall mean gene expression, which was twofold higher in RP genes (100.5±12.7) compared to GR genes (50.1 ±1.1), and fivefold higher compared to VG loci (20.3±2.6).

These findings are generally consistent with greater evolutionary constraint on loci with broader expression domains and, at least for broadly expressed (general) genes, greater magnitudes of gene expression.

3.6 Adaptively evolving reproductive proteins are not enriched for specific functional categories, but do include roles in sexual interactions

We found no evidence for enrichment of functional gene ontology categories among loci for which we found dN/dS>1, including among reproductive loci inferred to be adaptively evolving (data not shown). Nonetheless, among these reproductive loci, we did identify several genes with clear functional roles in sexual interactions (Table S12).

4 Discussion

Reproductive proteins appear to evolve rapidly in many animal groups. Here we evaluated evidence for elevated reproductive protein evolution in four closely-related flowering plant species, along with the potential influence of three factors on observed patterns: variation in mating system (and therefore the possible strength of sexual selection), in gametophytic gene expression (and therefore the strength of haploid selection), and in the breadth and/or magnitude of gene expression. We found evidence for elevated rates of reproductive protein evolution, globally across the genome and in some specific reproductive tissues. Moreover, these elevated rates appear to be more characteristic of female- rather than male-specific loci. Among the three broad factors that might influence these patterns, we found that more rapid evolution was consistently associated with more narrow domains of gene expression, but not with expression in haploid versus diploid tissues; evolutionary rate differences between lineages with different mating systems (self-compatible versus self-incompatible) were more complex and not associated with specific shifts in reproductive protein evolution. Here we discuss these findings in light of previously observed patterns of elevated reproductive protein evolution, and the factors proposed to drive these patterns.

4.1 Female reproductive proteins tend to evolve more rapidly

485 Our evidence for elevated rates of reproductive protein evolution was observed both as a
486 modest but significant increase in mean dN/dS across all reproductive loci, and as a significantly
487 larger proportion of loci showing evidence of adaptive divergence between species (dN/dS>1).
488 Interestingly, however, this elevated evolution appears to be more clearly associated with
489 proteins that have female- rather than male-specific functions. Indeed, female reproductive loci
490 predominate among our most rapidly evolving loci, both generally and in analyses of tissue-
491 specific loci. Of the tissue-specific loci found to have dN/dS>1, all are expressed in a female-
492 specific tissue (either style or ovule).
493
494 This finding is intriguing. Some previous studies have detected accelerated evolution in proteins
495 with specific female-associated functions, for example, egg proteins involved in sperm-egg
496 interactions (e.g. Galindo et al. 2003). Nonetheless, most studies that include both male and
497 female proteins have found mixed evidence for elevated female-biased or female-specific
498 evolution (e.g. Wong 2010, Haerty et al. 2007; including in plants, e.g. Gossman et al. 2016),
499 compared to much clearer accelerated evolution in male-specific or male-biased proteins
500 (reviewed in Wong 2011, Wilburn and Swanson 2016). Some of this male-female difference
501 could simply reflect a comparative lack of knowledge about female-associated proteins, so that
502 these might be systematically underexamined in many comparisons. Indeed, most prior studies
503 in animals (reviewed, for example, in Clark et al. 2006, Dapper and Wade 2020), as well as some
504 in plants (Arunkumar et al. 2013, Harrison et al. 2019, and references therein), have focused
505 largely or entirely on reproductive proteins with male-specific functions. Here, our loci were
506 defined by tissue-specific patterns of gene expression, rather than from *a priori* expectations of
507 function, which might have helped to ameliorate any such biases.
508
509 Our data also indicates a role for adaptive evolution in this accelerated female protein
510 divergence. Elevated rates of non-synonymous change (but values of dN/dS still less than 1)
511 could result from relaxed purifying selection (see further below) or because positive selection
512 has only acted on a subset of sites within each gene, or a combination of these effects.
513 However, we also found a larger proportion of female (style and ovule) loci with dN/dS>1,
514 suggesting a concentration of adaptive female protein evolution specifically involving post-
515 pollination reproductive processes. One driver of this could be sexual interactions. Unlike in
516 many animals, where female choice can be exerted both before and after mating (the latter as
517 'cryptic female choice') (Andersson and Simmons 2006), in angiosperms male-female (pollen-
518 pistil) interactions can only act after pollen arrives at a flower. Therefore all evolutionary
519 dynamics involving female choice or male-female interactions must act upon loci expressed in
520 this relatively narrow post-mating window. Indeed, among our rapidly evolving (dN/dS>1) style
521 loci, we detect at least one protein with putative functions in pollen interactions (a pistil-
522 specific extension-like protein; Solyc02g078080), in addition to several biosynthetic enzymes
523 (Table S12). (In contrast, 2/3 rapid ovule-specific genes have potential functions in stress
524 response, possibly suggesting roles that are not directly related to sexual interactions—see
525 below.) Moreover, in other angiosperms, 'female' loci that show rapid adaptive divergence are
526 also mostly pistil-specific proteins involved in selection among different pollen genotypes (via
527 roles in genetic self-incompatibility) (Clark et al. 2006), again suggesting the importance of post-
528 pollination sexual dynamics in accelerated reproductive protein evolution.

529
530 Whether this potential adaptive explanation extends to explain the general elevation of female
531 reproductive protein evolution (Figure 2) or reproductive proteins overall (Figure 1) is not yet
532 clear. One prior study that compared protein evolution in male, female, and non-reproductive
533 tissues in *Arabidopsis thaliana* and two relatives, also found that mean dN/dS was elevated in
534 loci with expression biased towards female tissues (egg cell, central cell, and synergid cells:
535 three haploid tissues within the female gametophyte) but not in pollen-biased proteins
536 (Gossman et al. 2014). This pattern was attributed to relaxed selection on female-biased loci,
537 partly because additional site-specific tests did not consistently detect evidence of positive
538 selection in these loci (Gossman et al. 2014). Because our four species are very closely related,
539 we do not have sufficient power to assess site-specific evidence for adaptive evolution in loci
540 whose estimated dN/dS is less than 1. Future analyses that pair divergence-based estimates
541 with polymorphism data within species—that can provide alternative and often more powerful
542 tests of positive and relaxed selection (e.g. Gossman et al. 2016, Arunkumar et al. 2013)—will
543 be helpful in differentiating the relative influence of these different forms of selection on the
544 broader patterns of elevated dN/dS detected here.
545
546 Regardless, if elevated female protein evolution is at least partly driven by post-pollination
547 sexual interactions, it is interesting that we do not also observe elevated evolution in pollen-
548 acting proteins. Of prior analyses in angiosperms, several have detected broadly elevated
549 evolution in pollen proteins (e.g., Arunkumar et al. 2013, Gossman et al. 2016, Harrison et al.
550 2019). Pollen protein evolution could be more constrained than style or ovule evolution due to
551 factors like differences in the degree of tissue-specificity or in the operation of haploid
552 selection, but these explanations are not generally consistent with our data (see further below).
553 Alternatively, it's possible that our analysis did not capture all the male proteins relevant to
554 post-pollination sexual interactions. For instance, pollen loci whose expression requires direct
555 interactions with the style and/or other pollen tubes would not be captured in our expression
556 atlas (which was generated from RNA expression in mature ungerminated and *in vitro*
557 germinated pollen; Table S1). There is evidence in other angiosperms that pollen
558 transcriptomes can be actively modified during their transit through the style, resulting in
559 changes in the suite of proteins being expressed (e.g., Qin et al. 2009, Mizukami et al. 2019).
560 Accounting for and including these genes would require transcriptomes from 'mixed' post-
561 pollination tissue, as well as specific approaches to differentiate pollen- from stylar-expression
562 in these mixed samples (e.g. Pease et al. 2016b). Importantly, this limitation would also extend
563 to any style- or ovule-specific proteins that are only elicited during pollination, so may not be a
564 complete explanation for the general difference we detect between male and female loci here.
565
566 Other explanations for the under-accounting of pollen loci could include technical effects such
567 as detection limits on gene expression in pollen, and/or exclusion of pollen-expressed loci that
568 are especially short or rapidly evolving, due to filtering during transcript-mapping steps. In
569 addition, by using dN/dS comparisons only, we would not have detected protein changes that
570 involve duplicate, chimeric, or novel genes, and/or if different loci are the targets of selection in
571 different lineages. If these latter factors do contribute to our failure to detect elevated pollen
572 protein evolution here, it suggests that different dynamics shape pollen protein identity and/or

evolution in *Solanum* compared to other angiosperms that have shown evidence for elevated evolution using metrics like dN/dS (e.g. Arunkumar et al. 2013, Gossman et al. 2016, Harrison et al. 2019). Regardless, with the data that we currently have, we do not see evidence that male functions are subject to especially intense sexual selection, particularly in comparison to female reproductive proteins.

4.2 Rates of protein evolution are affected by tissue-specificity but not haploid expression

Our results can also suggest factors that shape the patterns of reproductive protein evolution that we do observe, at least among three broad factors (gene expression level, haploid expression, and/or mating system variation) that we addressed here.

Of these, we find the strongest evidence that rates of protein evolution are influenced by the breadth and magnitude of gene expression. Many studies have shown higher rates of protein evolution in genes with more narrow expression domains and/or lower levels of gene expression (Meisel 2011, and references therein), a relationship that is thought to reflect lower constraint on tissue-specific proteins and stronger purifying selection on genes with very high expression levels or broad expression domains. We similarly found that higher rates of protein evolution were generally associated greater tissue-limitation and lower levels of gene expression in our species. However, we infer that tissue-limitation (a more narrow expression domain) is likely more important for shaping reproductive protein evolution specifically, as lower mean gene expression was not significantly associated with higher protein evolution for reproductive genes (Figure S1, Table 4), and reproductive loci also had much higher mean gene expression levels in comparison to generally-expressed genes and, especially, vegetative loci.

High tissue-specificity has been proposed as an important contributor to elevated rates of reproductive protein evolution—and a better predictor of this than sex-biased gene expression *per se*—in both animals (Meisel 2011) and some plants (Veltsos 2019). For example, in *Arabidopsis thaliana* and close relatives, Harrison et al. (2019) found faster protein evolution in both pollen-specific and tissue-specific sporophytic genes compared to loci expressed in >1 tissue, although pollen-specific loci still retain the highest evolutionary rates in this and other analyses in these species (e.g. Szovenyi et al. 2013, Gossman et al. 2016). Arunkumar et al. (2013) also inferred a role for tissue-limitation or specificity in elevated dN/dS in *Capsella grandiflora* pollen proteins (although this elevation was attributed to different forms of selection—relaxed purifying selection in pollen germ cell genes versus greater purifying and positive selection in pollen tube loci—depending on the specific sub-tissues within pollen). These prior inferences, and our observations, suggest that higher tissue-specificity could facilitate broadly elevated rates of protein evolution. Nonetheless, tissue-specificity cannot provide a complete explanation of elevated reproductive protein evolution in our dataset, as tissue-specific non-reproductive proteins still evolve more slowly than tissue-specific (female) reproductive proteins (Table 4).

In comparison, we found no rate differences between loci expressed in haploid versus diploid tissues, and therefore little evidence for a strong effect of haploid selection in driving

differences between reproductive and non-reproductive proteins. Haploid selection has two projected effects on haploid-expressed loci—increasing the efficacy of positive selection on advantageous alleles, and of purifying selection on deleterious alleles (Immler and Otto 2018, Mattila et al. 2018)—compared to diploid-expressed loci. We have limited ability to differentiate these two effects with our particular estimate of protein evolution, but our observations are not consistent with a strong effect of either. In terms of accelerating adaptive protein evolution, we found no significant elevation of evolutionary rates in loci limited to gametophytic-tissues in general, and in ovule-limited and pollen-limited loci specifically (Table 3). Similarly, we find little evidence for stronger or more widespread purifying selection on haploid-expressed loci, which would be observed as lower mean rates of protein evolution in gametophytic loci. Stronger purifying selection due to haploid expression has been proposed as one reason why sex-biased loci (generally, pollen expressed loci) do not appear to evolve faster than non-sex-biased loci in several dioecious species (Sanderson et al. 2019, Cossard et al. 2019; discussed in Muyle 2019, but see also Veltsos 2019) and, indeed, sometimes appear to evolve more slowly (e.g., Darolti et al. 2018). Our data do not suggest stronger purifying selection in either gametophytic-limited proteins in general, or ovule- and pollen-limited loci specifically, in our species. However, because we have relatively few pollen-specific loci (discussed above) we cannot exclude the possibility that intermediate rates of pollen protein evolution result from a tension between both positive and purifying selection predicted to act on these loci. Nonetheless, our present data indicates little evidence for a global effect of haploid gene expression on molecular evolutionary rates.

4.3 Mating system variation does not does not differentially influence reproductive protein evolution

Our results also did not support a strong consistent effect of mating system differences—and therefore predicted differences in the strength of sexual selection—on patterns of reproductive protein evolution. Interpreting these comparisons is complex because differences in genetic effective population size are expected to have global effects on rates of molecular evolution (Charlesworth and Wright 2001, and see Introduction) that are unrelated to sexual selection *per se*. The broad patterns of dN/dS we observed generally reflected these expected global effects of differences in Ne due to mating system variation: in the SC lineages, mean dN/dS was higher in both RP and GR genes, consistent with a general relaxation of selection in more selfing lineages (Charlesworth and Wright 2001) regardless of whether or not loci were associated with reproductive functions (Table 4). This genome-wide effect might have overwhelmed more subtle effects of mating system variation acting specifically on loci involved in male-male competition and/or female choice. In addition, the transition to self-compatibility is relatively recent (less than 1 MY) among our species, and self-compatible lineages still retain significant potential for outcrossing (Rick 1979, Vosters et al. 2014). These factors mean there may be relatively modest differences in the historical and current opportunity for sexual selection among our lineages, and therefore little opportunity to observe systematic consequences of this transition as differential changes in evolutionary rates. (Statistically, this is compounded in our particular dataset by the two SC species branches being substantially shorter than those for our two SI species.) This contrasts with other angiosperms, such as *A. thaliana*, where the

661 transition to self-compatibility is about as old but the consequent effects on outcrossing rates
662 has been much more severe (selfing rates can be >95%; Harrison et al. 2019, and references
663 therein).
664
665 Interestingly, several previous analyses in animal systems have similarly found equivocal
666 evidence for differential effects of mating system variation on the evolution of reproductive
667 tract proteins (reviewed in Wong 2011). For example, in studies of primates, evolutionary rates
668 in specific sperm and copulatory plug proteins are positively associated with common proxies
669 for sexual selection, such as residual testis size and estimated number of mates (e.g. Dorus et
670 al. 2004, Martin-Coello et al. 2009). However other studies find that phylogeny-wide signals of
671 positive selection in reproductive tract proteins are not associated specifically with polyandrous
672 (as opposed to monogamous) lineages (e.g., Ramm et al. 2008, Finn and Civetta 2010; and see
673 Table 1 in Wong 2011). Limitations on phylogenetic models that can assess associations
674 between molecular evolutionary rates and changes in phenotypic characters, and variation
675 among lineages in the specific targets of sexual selection (Wong 2011), might explain relatively
676 weak effects of mating system variation on protein evolution rates, despite the importance of
677 sexual selection in driving faster reproductive protein evolution. Adding more species, and
678 focusing on pairs of SI-SC sister species, might help in addressing some of these limitations in
679 future analyses within *Solanum*.
680
681 Alternatively, the weak effect of mating system variation—both here and in prior studies—
682 might be because the predominant cause of globally elevated reproductive protein evolution is
683 not sexual selection. For example, pathogen resistance responses are also known to be
684 associated with elevated rates of protein evolution—driven by antagonistic host-pathogen
685 evolution—and might be concentrated among reproductive proteins if mating often leads to
686 the exchange of pathogens (Clark et al. 2006, Wong 2011). In principle, the identity of loci
687 inferred to be under positive selection could be helpful in differentiating among alternative
688 selective agents responsible for driving this adaptive divergence. Of the genes in our dataset
689 that show elevated evolution, we cannot yet draw strong conclusions about whether they
690 better reflect roles related to sexual interactions versus responses to non-sexual environmental
691 factors; many of them have unknown functions or general functions that do not *a priori* provide
692 specific support for either inference (Table S12). We do detect at least one style-specific protein
693 with a clear functional connection to pollen-pistil interactions, however our rapidly-evolving
694 ovule-specific genes are implicated in stress responses, which might indicate a greater role for
695 natural rather than sexual selection in shaping adaptive responses in this class of loci.
696
697 **4.4 Consequences of rapid protein evolution for the evolution of reproductive differences**
698 **and speciation**
699
700 Finally, among the motivations for examining patterns of reproductive protein evolution is to
701 understand their possible role in lineage diversification. If reproductive proteins do indeed
702 evolve more rapidly than other classes of protein, then perhaps they play an outsized role in
703 the emergence of functional differences and reproductive isolating barriers between lineages.
704 For example, evidence suggests post-mating prezygotic traits frequently evolve rapidly in

705 animals and, when divergent between species, could contribute to reproductive isolation
706 (Howard 1999, Snook et al. 2009, Palumbi 2009, Castillo and Moyle 2014). Similarly, divergence
707 in post-pollination prezygotic reproductive molecules could be important contributors to
708 species barrier formation among angiosperms (Moyle et al. 2014). Here we did detect evidence
709 for a global elevation of protein evolution in tissues that mediate these post-pollination
710 prezygotic stages, especially of female-specific loci. Our observations are consistent with a role
711 for tissue specificity in facilitating these elevated rates, but not for haploid selection or, as yet,
712 for differences in selective dynamics that are associated with recent shifts in mating system.
713 The evolutionary factors that might drive this more rapid evolution therefore remain to be
714 clarified. Regardless, our results indicate that rapid reproductive divergence is especially
715 characteristic of proteins with female-specific functions among our species. These specific
716 functions can be critical for maintaining coordinated sexual signals and ensuring fertilization
717 post-pollination (Bernasconi et al. 2001, Swanson et al. 2004, Moyle et al. 2014). Therefore
718 their rapid divergence might be particularly influential in the initial stages of reproductive
719 isolation and, thereby, in speciation, in this highly diverse plant genus.
720
721
722 **5 Author contributions**
723
724 LCM, MW, and MJSG designed the study. MW and MJSG performed the analyses. LCM wrote
725 the manuscript with assistance from MW and MJSG.
726
727
728 **6 Funding**
729
730 This work was supported by NSF grant IOS-1127059 and NSF grant DEB-1856469 to LCM.
731
732
733 **7 Acknowledgments**
734
735 Thanks to CJ Jewell for collecting ovule tissue, Christie Bergeon for RNA extractions, and the
736 Indiana University Center for Genomics and Bioinformatics for assistance with library
737 construction and sequencing.
738

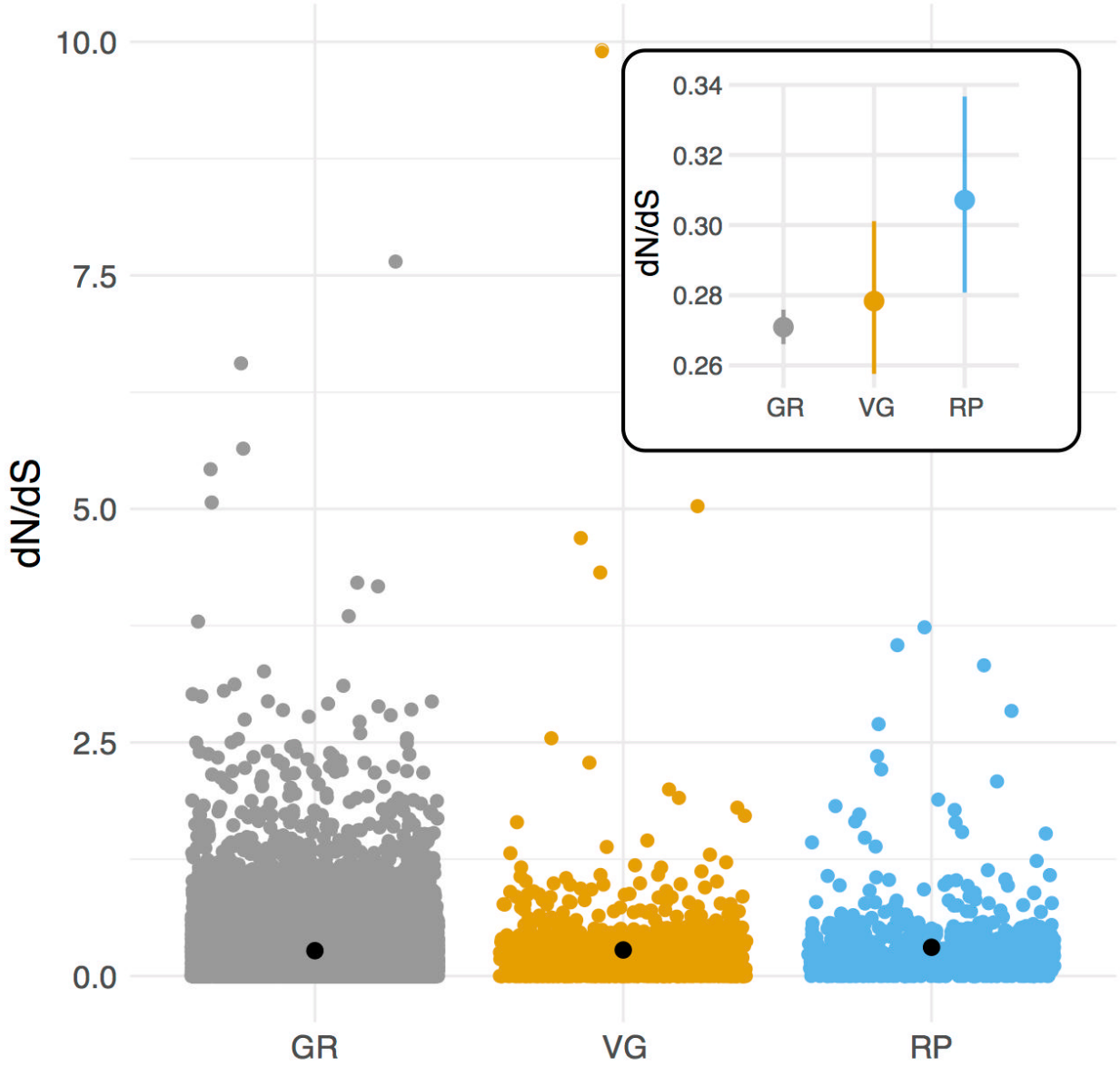


Figure 1: Per locus protein evolution (dN/dS) among loci expressed in reproductive (RP), vegetative (VG), and both reproductive and non-reproductive (general/GR) tissues in Solanum species. Black circles indicate means for each group. Inset: Estimated mean (circle) and 95% CI of dN/dS for each group of loci.

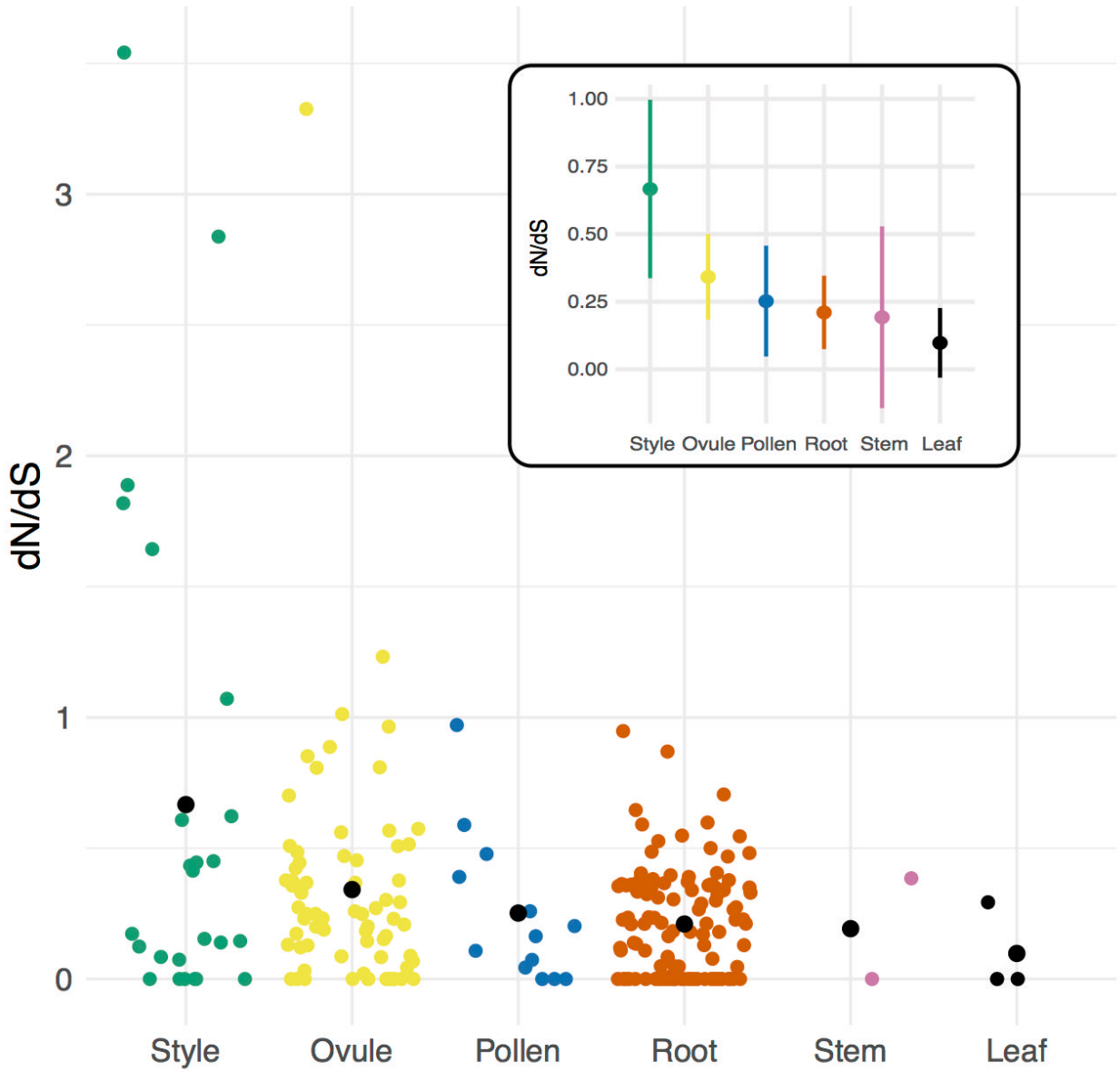


Figure 2: Per locus protein evolution (dN/dS) among groups of tissue-specific loci in Solanum species. Black circles indicate means for each group. Inset: Estimated mean (circle) and 95% CI of dN/dS for each group of loci.

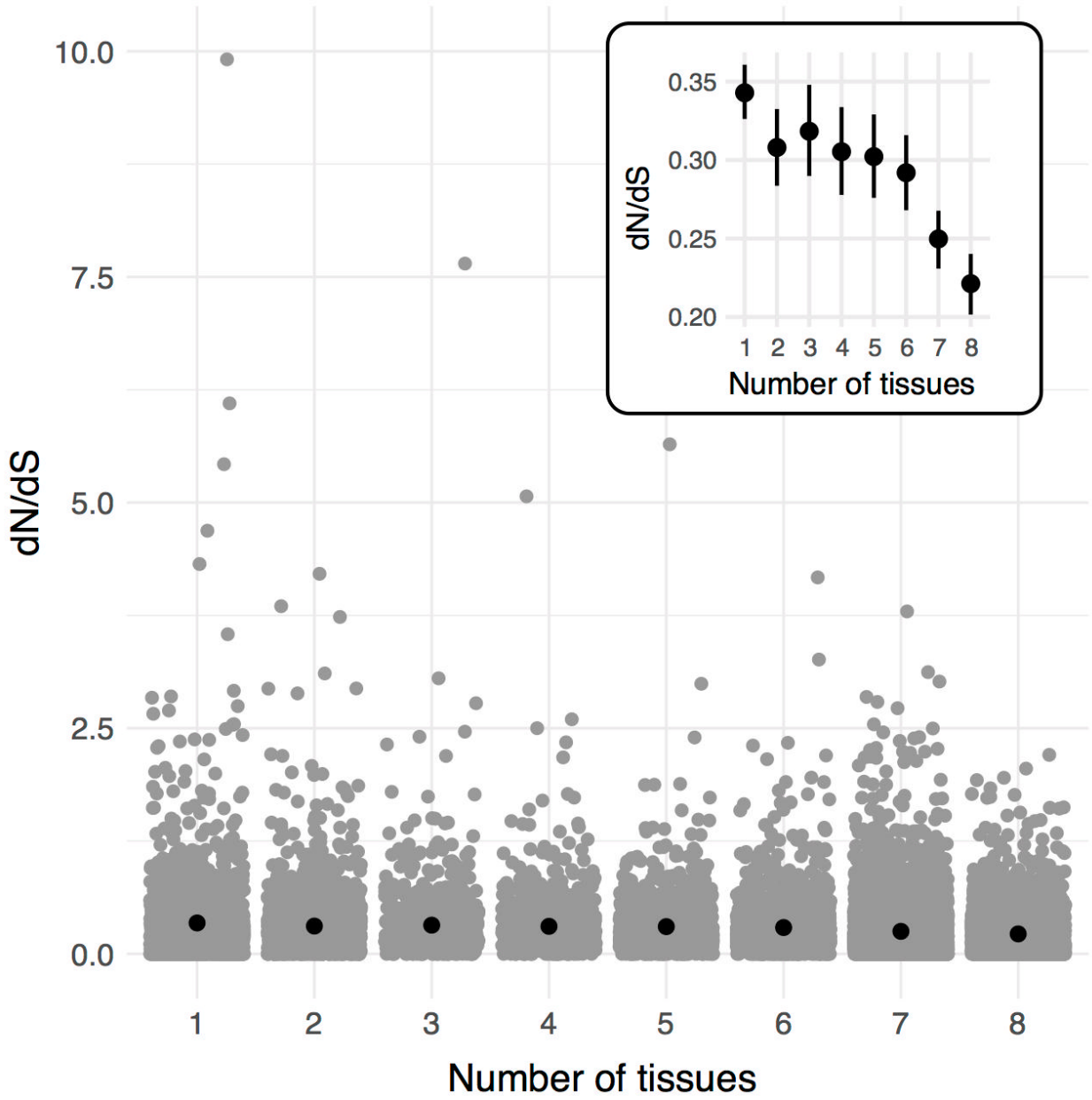
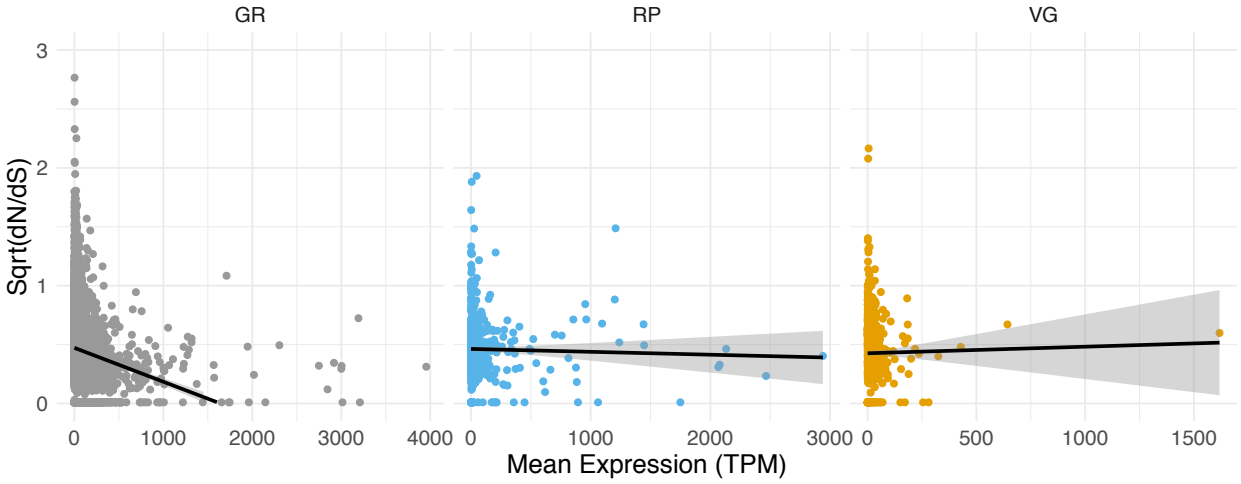


Figure 3: Per locus protein evolution (dN/dS) among groups of loci expressed (>2 TPM) in different numbers of tissues (from 1 to 8) in Solanum species. Black circles indicate means for each group. Inset: Estimated mean (circle) and 95% CI of dN/dS for each group of loci.



Supplementary Figure 1: Relationship between per locus mean level of gene expression (TPM) and per locus protein evolution (dN/dS) in loci expressed in (left) general/GR, (middle) reproductive/RP, and (right) vegetative/VG tissues in Solanum species. Each point indicates an individual locus. Regression lines shown in black.

770 **Tables**

771

772 **Table 1. Rates of protein evolution (dN/dS) in different broad classes of loci.** dN/dS calculated
773 for each locus using a one ratio test in PAML.

774

		Per locus dN/dS				Loci with dN/dS>1		
Class	N loci	mean	SE	Median	Tukey test[1]	Number	Proportion2	Fisher's exact tests[1]
General	16221	0.2709	0.0026	0.20	a	399	0.0246	a
Vegetative	923	0.2783	0.0111	0.19	ab	26	0.0282	ab
Reproductive	670	0.3071	0.0142	0.21	b	28	0.0418	b

775

776 1. see Table S3.
777 2. The proportion of genes with dN/dS>1 differed among classes of loci (X^2=7.9696, p=0.0186).

778

779

780 **Table 2: Rates of protein evolution (dN/dS) in tissue-specific loci**

781

	Per locus dN/dS					Loci with dN/dS>1	
Group	N	Mean	SE	Median	Tukey test[1]	Number	Proportion[2]
Leaf	3	0.0978	0.0657	0.00	a	0	0
Stem	2	0.1925	0.1714	0.19	abc	0	0
Root	124	0.2103	0.0693	0.19	ab	0	0
Pollen	13	0.2521	0.1046	0.16	abc	0	0
Ovule	71	0.3418	0.0809	0.25	bc	3	0.042
Style	25	0.6668	0.1685	0.17	c	6	0.240

782

783 1. see Table S6.
784 2. The proportion of genes with dN/dS>1 differed by tissue (X^2= 33.707, p=2.72 x 10-6). One-
785 sided Fisher's exact tests indicated this proportion was greater in style compared to all other
786 tissues (p=5.046 x 10-5), and greater in ovule compared to all other tissues except style
787 (p=0.03756).

788 **Table 3: Rates of protein evolution (dN/dS) in gametophytic (GM) versus sporophytic (SP) or**
789 **gametophytic+sporophytic (GM+SP) loci.** Within each GLM, tissue groups did not differ in
790 mean dN/dS (Table S7).
791

	dN/dS				Loci with dN/dS>1		
Group	N	mean	SE	Median	Number	Proportion	p[1]
a) Gametophytic vs Sporophytic							
Gametophytic only	86	0.3275	0.0576	0.24	3	0.0349	0.5574
Sporophytic only	2033	0.2958	0.0586	0.21	68	0.0334	
b) Gametophytic vs GM+SP							
Gametophytic only	86	0.3275	0.0393	0.24	3	0.0349	0.2833
Gametophytic + sporophytic	15778	0.2685	0.0394	0.20	383	0.0243	
c) Ovule loci							
Ovule only	71	0.3418	0.0036	0.25	3	0.0423	0.8542
Ovule + sporophytic	7884	0.2874	0.0448	0.21	225	0.0285	
d) Pollen loci							
Pollen only	13	0.2521	0.0679	0.16	0	0.0000	0.8316
Pollen + sporophytic	277	0.2496	0.0695	0.18	4	0.0144	

792
793 1. Fisher's exact tests
794
795
796 **Table 4: Estimates of dN/dS for SC and SI branches separately, for loci concatenated within**
797 **each broad class of genes.** Results are from 3 ratio model in PAML, with input species tree:
798 ((Slyc #1, Spim #1), (Spen #2, Shab #2)).
799

Class	# genes	# codons	Null model	Variable rate model	2*LnL	p value	SI-branch dN/dS	SC-branch dN/dS	Internal Branch	SC-S diff
GR	16221	6619178	-28576265.18	-28575900.38	729.61282	3.686 10x-159	0.2224	0.2617	0.1901	0.03
RP	904	110171	-478937.9344	-478932.4222	11.024336	0.00404	0.2285	0.2623	0.1999	0.03
VE	1300	78355	-344551.0442	-344546.389	9.310318	0.00951	0.2267	0.2042	0.1850	-0.02

800

801 **Table 5: Rates of protein evolution (dN/dS) in loci according to breadth of tissue expression.**
802

# Tissues expressed in	N loci	Mean sqrt(dN/dS)	SE	Median	Posthoc Tukey test[1]
1	1892	0.3429	0.0088	0.24	d
2	1528	0.3080	0.0125	0.22	cd
3	901	0.3183	0.0148	0.23	cd
4	930	0.3053	0.0142	0.23	cd
5	1083	0.3022	0.0135	0.24	c
6	1507	0.2919	0.0122	0.21	c
7	7206	0.2498	0.0094	0.19	b
8	3173	0.2213	0.0098	0.16	a

803
804 1. see Table S10
805
806
807 **Table 6: Relationship between mean gene expression level and estimated dN/dS, across loci**
808 **in each broad class (GR, RP, VG) of loci (Linear model)**
809

	DF	SS	MS	F	Pr > F
Expression level	1	15.44	15.4355	243.85	2.2 x 10-16
Class	2	1	0.4987	7.88	0.00038
Expression level x Class	2	3.34	1.6689	26.36	3.72 x 10-12
	15086	954.94	0.0633		

810
811

812 **8 References**

813 ANDERSSON M, SIMMONS LW. 2006 Sexual selection and mate choice. Trends in Ecology & Evolution.
814 21(6):296-302.

815 ANDERSSON MB. 1994 *Sexual Selection*. Princeton University Press, Princeton, NJ.

816 ARUNKUMAR R, JOSEPHS EB, WILLIAMSON RJ, WRIGHT SI. 2013 Pollen-specific, but not sperm-specific,
817 genes show stronger purifying selection and higher rates of positive selection than sporophytic genes
818 in *Capsella grandiflora*. Mol Biol Evol. 30(11):2475-2486.

819 ASHBURNER M, BALL CA, BLAKE JA, BOTSTEIN D, BUTLER H, CHERRY JM, DAVIS AP, DOLINSKI K,
820 DWIGHT SS, EPPIG JT, HARRIS MA, HILL DP, ISSEL-TARVER L, KASARSKIS A, LEWIS S, MATESE JC,
821 RICHARDSON JE, RINGWALD M, RUBIN GM, SHERLOCK G, GENE ONTOLOGY C. 2000 Gene Ontology:
822 tool for the unification of biology. Nature Genetics. 25(1):25-29.

823 BERNASCONI G, ASHMAN TL, BIRKHEAD TR, BISHOP JDD, GROSSNIKLAUS U, KUBLI E, MARSHALL
824 DL, SCHMID B, SKOGSMYR I, SNOOK RR, TAYLOR D, TILL-BOTTRAUD I, WARD PI, ZEH DW,
825 HELLRIEGEL B. 2004 Evolutionary ecology of the prezygotic stage. Science. 303(5660):971-975.

826 BOLGER AM, LOHSE M, USADEL B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence
827 data. Bioinformatics. 30(15):2114-2120.

828 BRANDVAIN Y, HAIG D. 2005 Divergent mating systems and parental conflict as a barrier to
829 hybridization in flowering plants. American Naturalist. 166(3):330-338.

830 CASTILLO DM, MOYLE LC. 2014 Intraspecific sperm competition genes enforce post-mating
831 species barriers in Drosophila. Proceedings of the Royal Society B-Biological Sciences.
832 281(1797).

833 CHARLESWORTH B, MORGAN MT, CHARLESWORTH D. 1993 The effect of deleterious mutations on
834 neutral molecular variation Genetics. 134(4):1289-1303.

835 CHARLESWORTH D, CHARLESWORTH B. 1992 The effects of selection in the gametophytic stage on
836 mutational load Evolution. 46(3):703-720.

837 CHARLESWORTH D, WRIGHT SI. 2001 Breeding systems and genome evolution. Current Opinion in
838 Genetics & Development. 11(6):685-690.

839 CLARK NL, AAGAARD JE, SWANSON WJ. 2006 Evolution of reproductive proteins from animals and
840 plants. Reproduction. 131(1):11-22.

841 CONSORTIUM TGO. 2019 The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids
842 Research. 47(D1):D330-D338.

843 COSSARD GG, TOUPS MA, PANNELL JR. 2019 Sexual dimorphism and rapid turnover in gene expression
844 in pre-reproductive seedlings of a dioecious herb. Annals of Botany. 123(7):1119-1131.

845 CRUZ-GARCIA F, HANCOCK CN, MCCLURE B. 2003 S-RNase complexes and pollen rejection. Journal of
846 Experimental Botany. 54(380):123-130.

847 DAPPER AL, WADE MJ. 2020 Relaxed Selection and the Rapid Evolution of Reproductive Genes. Trends
848 in Genetics. 36(9):640-649.

849 DAROLTI I, WRIGHT AE, PUCHOLT P, BERLIN S, MANK JE. 2018 Slow evolution of sex-biased genes in the
850 reproductive tissue of the dioecious plant Salix viminalis. Molecular Ecology. 27(3):694-708.

851 DELPH LF, ASHMAN TL. 2006 Trait selection in flowering plants: how does sexual selection contribute?
852 Integrative and Comparative Biology. 46(4):465-472.

853 DELPH LF, HAVENS K. 1998 Pollen competition in flowering plants. In: *Sperm competition and sexual*
854 *selection*. (TR B, AP M, eds), pp. 149-173. Academic Press, San Diego, CA.

855 DOBIN A, DAVIS CA, SCHLESINGER F, DRENKOW J, ZALESKI C, JHA S, BATUT P, CHAISSON M,
856 GINGERAS TR. 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15-21.

857 DORUS S, EVANS PD, WYCKOFF GJ, CHOI SS, LAHN BT. 2004 Rate of molecular evolution of the seminal
858 protein gene SEMG2 correlates with levels of female promiscuity. Nature Genetics. 36(12):1326-1329.

859 FINN S, CIVETTA A. 2010 Sexual Selection and the Molecular Evolution of ADAM Proteins. Journal of
860 Molecular Evolution. 71(3):231-240.

861 GALINDO BE, VACQUIER VD, SWANSON WJ. 2003 Positive selection in the egg receptor for abalone
862 sperm lysin. Proceedings of the National Academy of Sciences of the United States of America.
863 100(8):4639-4643.

864 GOLDBERG EE, KOHN JR, LANDE R, ROBERTSON KA, SMITH SA, IGIC B. 2010 Species selection maintains
865 self-incompatibility. Science. 330(6003):493-495.

866 GOSSMANN TI, SALEH D, SCHMID MW, SPENCE MA, SCHMID KJ. 2016 Transcriptomes of Plant
867 Gametophytes Have a Higher Proportion of Rapidly Evolving and Young Genes than Sporophytes.
868 Molecular Biology and Evolution. 33(7):1669-1678.

869 GOSSMANN TI, SCHMID MW, GROSSNIKLAUS U, SCHMID KJ. 2014 Selection-Driven Evolution of Sex-
870 Biased Genes Is Consistent with Sexual Selection in Arabidopsis thaliana. Molecular Biology and
871 Evolution. 31(3):574-583.

872 HAERTY W, JAGADEESHAN S, KULATHINAL RJ, WONG A, RAM KR, SIROT LK, LEVESQUE L, ARTIERI CG,
873 WOLFNER MF, CIVETTA A, SINGH RS. 2007 Evolution in the fast lane: Rapidly evolving sex-related genes
874 in drosophila. Genetics. 177(3):1321-1335.

875 HAFIDH S, FILA J, HONYS D. 2016 Male gametophyte development and function in angiosperms: a
876 general concept. Plant Reproduction. 29(1-2):31-51.

877 HARRISON MC, MALLON EB, TWELL D, HAMMOND RL. 2019 Deleterious Mutation Accumulation in
878 Arabidopsis thaliana Pollen Genes: A Role for a Recent Relaxation of Selection. Genome Biology and
879 Evolution. 11(7):1939-1951.

880 HOWARD DJ. 1999 Conspecific sperm and pollen precedence and speciation. Annual Review of
881 Ecology and Systematics. 30:109-132.

882 IMMLER S. 2019 Haploid Selection in "Diploid" Organisms. In: *Annual Review of Ecology, Evolution, and*
883 *Systematics, Vol 50*, Vol. 50: *Annual Review of Ecology Evolution and Systematics* (Futuyma DJ, ed), pp.
884 219-236.

885 IMMLER S, OTTO SP. 2018 The evolutionary consequences of selection at the haploid gametic stage.
886 American Naturalist. 192(2):241-249.

887 ICHIHASHI Y, AGUILAR-MARTINEZ JA, FARHI M, CHITWOOD DH, KUMAR R, MILLON LV, PENG J,
888 MALOOF JN, SINHA NR. 2014 Evolutionary developmental transcriptomics reveals a gene
889 network module regulating interspecific diversity in plant leaf shape. Proceedings of the
890 National Academy of Sciences of the United States of America. 111(25):E2616-E2621.

891 KOENIG D, JIMENEZ-GOMEZ JM, KIMURA S, FULOP D, CHITWOOD DH, HEADLAND LR, KUMAR R,
892 COVINGTON MF, DEVISETTY UK, TAT AV, TOHGE T, BOLGER A, SCHNEEBERGER K, OSSOWSKI S,
893 LANZ C, XIONG GY, TAYLOR-TEEPLES M, BRADY SM, PAULY M, WEIGEL D, USADEL B, FERNIE AR,
894 PENG J, SINHA NR, MALOOF JN. 2013 Comparative transcriptomics reveals patterns of selection
895 in domesticated and wild tomato. Proceedings of the National Academy of Sciences of the
896 United States of America. 110(28):E2655-E2662.

897 LANKINEN A, HYDBOM S, STRANDH M. 2017 Sexually antagonistic evolution caused by male-male
898 competition in the pistil. Evolution. 71(10):2359-2369.

899 LIAO Y, SMYTH GK, SHI W. 2014 featureCounts: an efficient general purpose program for
900 assigning sequence reads to genomic features. Bioinformatics. 30(7):923-930.

901 LLOYD DG, WEBB CJ. 1977 Secondary sexual characters in plants Botanical Review. 43(2):177-216.

902 MARTIN-COELLO J, DOPAZO H, ARBIZA L, AUSIO J, ROLDAN ERS, GOMENDIO M. 2009 Sexual selection
903 drives weak positive selection in protamine genes and high promoter divergence, enhancing sperm
904 competitiveness. Proceedings of the Royal Society B-Biological Sciences. 276(1666):2427-2436.

905 MATTILA TM, LAENEN B, SLOTTE T. 2018 Population genomics of transitions to selfing in Brassicaceae
906 model systems. Arxiv. http://arxiv.org/abs/1809.04872.

907 MAZER SJ, HENDRICKSON BT, CHELLEW JP, KIM LJ, LIU JW, SHU J, SHARMA MV. 2018 Divergence in
908 pollen performance between Clarkia sister species with contrasting mating systems supports
909 predictions of sexual selection. Evolution. 72(3):453-472.

910 MAZER SJ, HOVE AA, MILLER BS, BARBET-MASSIN M. 2010 The joint evolution of mating system and
911 pollen performance: Predictions regarding male gametophytic evolution in selfers vs. outcrossers.
912 Perspectives in Plant Ecology Evolution and Systematics. 12(1):31-41.

913 MCCLURE B, CRUZ-GARCIA F, ROMERO C. 2011 Compatibility and incompatibility in S-RNase-based
914 systems. Annals of Botany. 108(4):647-658.

915 MEISEL RP. 2011 Towards a More Nuanced Understanding of the Relationship between Sex-Biased
916 Gene Expression and Rates of Protein-Coding Sequence Evolution. Molecular Biology and Evolution.
917 28(6):1893-1900.

918 MI H, HUANG X, MURUGANUJAN A, TANG H, MILLS C, KANG D, THOMAS PD. 2019 PANTHER version
919 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. . Nucleic
920 Acids Research. 47(D1):D419-D426.

921 MIZUKAMI AG, INATSUGI R, JIAO J, KOTAKE T, KUWATA K, OOTANI K, OKUDA S, SANKARANARAYANAN
922 S, SATO Y, MARUYAMA D, IWAI H, GARENAUX E, SATO C, KITAJIMA K, TSUMURAYA Y, MORI H,
923 YAMAGUCHI J, ITAMI K, SASAKI N, HIGASHIYAMA T. 2016 The AMOR Arabinogalactan Sugar Chain
924 Induces Pollen-Tube Competency to Respond to Ovular Guidance. Current Biology. 26(8):1091-1097.

925 MIZUTA Y, HIGASHIYAMA T. 2018 Chemical signaling for pollen tube guidance at a glance. Journal of
926 Cell Science. 131(2).

927 MOORE JC, PANNELL JR. 2011 Sexual selection in plants. Current Biology. 21(5):R176-R182.

928 MOYLE LC, JEWELL CP, KOSTYUN JL. 2014 Fertile approaches to dissecting mechanisms of premating
929 and postmating prezygotic reproductive isolation. Current Opinion in Plant Biology. 18:16-23.

930 MUYLE A. 2019 How different is the evolution of sex-biased gene expression between plants and
931 animals? A commentary on: 'Sexual dimorphism and rapid turnover in gene expression in
932 prereproductive seedlings of a dioecious herb'. Annals of Botany. 123(7):IV-V.

933 PEASE JB, HAAK DC, HAHN MW, MOYLE LC. 2016A Phylogenomics reveals three sources of adaptive
934 variation during a rapid radiation. PLoS Biology. 14(2):e1002379.

935 PEASE JB, GUERRERO RF, SHERMAN NA, HAHN MW, MOYLE LC. 2016B Molecular mechanisms of
936 postmating prezygotic reproductive isolation uncovered by transcriptome analysis. Molecular Ecology.
937 25(11):2592-2608.

938 PETERS MAE, WEIS AE. 2018 Selection for pollen competitive ability in mixed-mating systems.
939 Evolution. 72(11):2513-2536.

940 QIN Y, LEYDON AR, MANZIELLO A, PANDEY R, MOUNT D, DENIC S, VASIC B, JOHNSON MA, PALANIVELU
941 R. 2009 Penetration of the Stigma and Style Elicits a Novel Transcriptome in Pollen Tubes, Pointing to
942 Genes Critical for Growth in a Pistil. Plos Genetics. 5(8).

943 R CORE TEAM. 2015. R: A language and environment for statistical computing. Vienna, Austria:
944 Sponsored by the R Foundation for Statistical Computing.

945 RAMM SA, OLIVER PL, PONTING CP, STOCKLEY P, EMES RD. 2008 Sexual selection and the adaptive
946 evolution of mammalian ejaculate proteins. Molecular Biology and Evolution. 25(1):207-219.

947 RICK CM. 1979 Biosystematic studies in Lycopersicon and closely related species of Solanum. In: *The*
948 *biology and taxonomy of the Solanaceae*. (Hawkes JG, Lester RN, Skelding AD, eds), pp. 667-679.
949 Academic Press Inc., New York.

950 RITCHIE MG. 2007 Sexual selection and speciation. Annual Review of Ecology, Evolution, and
951 Systematics. 38:79-102.

952 SANDERSON BJ, WANG L, TIFFIN P, WU ZQ, OLSON MS. 2019 Sex-biased gene expression in flowers,
953 but not leaves, reveals secondary sexual dimorphism in Populus balsamifera. New Phytologist.
954 221(1):527-539.

955 SARKINEN T, BOHS L, OLMSTEAD RG, KNAPP S. 2013 A phylogenetic framework for evolutionary study
956 of the nightshades (Solanaceae): a dated 1000-tip tree. Bmc Evolutionary Biology. 13.

957 SICARD A, LENHARD M. 2011 The selfing syndrome: a model for studying the genetic and evolutionary
958 basis of morphological adaptation in plants. Annals of Botany. 107(9):1433-1443.

959 SKOGSMYR I, LANKINEN A. 2002 Sexual selection: an evolutionary force in plants. Biological Reviews.
960 77(4):537-562.

961 SLOTTE T, BATAILLON T, HANSEN TT, ST ONGE KR, WRIGHT SI, SCHIERUP MH. 2011 Genomic

962 Determinants of Protein Evolution and Polymorphism in Arabidopsis. Genome Biology and Evolution.
963 3:1210-1219.

964 SNOOK RR, CHAPMAN T, MOORE PJ, WEDELL N, CRUDGINGTON HS. 2009 Interactions between
965 the sexes: new perspectives on sexual selection and reproductive isolation. Evolutionary
966 Ecology. 23(1):71-91.

967 STEPHENSON AG, BERTIN RI. 1983 Male competition, female choice, and sexual selection in plants. In:
968 *Pollination Biology*. (L R, ed), pp. 109-151. Academic Press, New York, NY.

969 STONE JL. 2002 Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility.
970 Quarterly Review of Biology. 77(1):17-32.

971 SWANSON R, EDLUND AF, PREUSS D. 2004 Species specificity in pollen-pistil interactions.
972 Annual Review of Genetics. 38:793-818.

973 SWANSON WJ, VACQUIER VD. 2002 The rapid evolution of reproductive proteins. Nature Reviews
974 Genetics. 3(2):137-144.

975 SWANSON WJ, VACQUIER VD. 2002 Reproductive protein evolution. Annual Review of Ecology and
976 Systematics. 33:161-179.

977 SZOVENYI P, RICCA M, HOCK Z, SHAW JA, SHIMIZU KK, WAGNER A. 2013 Selection Is No More Efficient
978 in Haploid than in Diploid Life Stages of an Angiosperm and a Moss. Molecular Biology and Evolution.
979 30(8):1929-1939.

980 TAKAYAMA S, ISOGAI A. 2005 Self-incompatibility in plants. Annual Review of Plant Biology. 56:467-
981 489.

982 THE GENE ONTOLOGY CONSORTIUM. 2019 The Gene Ontology Resource: 20 years and still GOing
983 strong. Nucleic Acids Research. 47(D1):D330-D338.

984 TURNER LM, HOEKSTRA HE. 2008 Causes and consequences of the evolution of reproductive proteins.
985 International Journal of Developmental Biology. 52(5-6):769-780.

986 VACQUIER VD, SWANSON WJ. 2011 Selection in the Rapid Evolution of Gamete Recognition Proteins in
987 Marine Invertebrates. Cold Spring Harbor Perspectives in Biology. 3(11).

988 VELTSOS P. 2019 Not all sex-biased genes are the same. New Phytologist. 221(1):10-11.

989 VOSTERS SL, JEWELL CP, SHERMAN NA, EINTERZ F, BLACKMAN BK, MOYLE LC. 2014 The timing of
990 molecular and morphological changes underlying reproductive transitions in wild tomatoes (Solanum
991 sect. Lycopersicon). Molecular Ecology. 23(8):1965-1978.

992 WALSH NE, CHARLESWORTH D. 1992 Evolutionary interpretations of differences in pollen-tube growth
993 rates Quarterly Review of Biology. 67(1):19-37.

994 WILBURN DB, SWANSON WJ. 2016 From molecules to mating: Rapid evolution and biochemical studies
995 of reproductive proteins. Journal of Proteomics. 135:12-25.

996 WILLSON MF. 1979 Sexual selection in plants American Naturalist. 113(6):777-790.

997 WILLSON MF. 1994 Sexual selection in plants - perspective and overview American Naturalist. 144:S13-
998 S39.

999 WONG A. 2010 Testing the effects of mating system variation on rates of molecular evolution in
1000 primates. Evolution. 64(9):2779-2785.

1001 WONG A. 2011 The molecular evolution of animal reproductive tract proteins: what have we learned
1002 from mating-system comparisons? International Journal of Evolutionary Biology. 2011(908735):1-9.

1003 YANG Z. 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution.
1004 24(8):1586-1591.
1005
1006