[Original Research Article]

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 also 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, some specific patterns of 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 observed 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 existing 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 have 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 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 patterns of selection across the genome, regardless of whether genes have reproductive functions. Because selfing reduces 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 current studies of reproductive protein evolution in plants.

133 In this study, we examined genome-wide patterns and rates of protein evolution among four
134 closely-related species from the diverse plant genus *Solanum*. Our aims were to compare
135 patterns of molecular evolution in loci expressed in different classes of tissue in order to
136 evaluate whether: a) reproductive-specific proteins evolve more rapidly, on average, than non-
137 reproductive proteins; b) 'male'-specific loci evolve at different rates than 'female'-specific loci;
138 c) genes expressed exclusively in gametophytic tissues evolve differently from genes expressed
139 in sporophytic (diploid) tissue or in both tissue types; and d) mating system variation (a
140 potential proxy for the expected strength of sexual selection and/or sexual conflict) affects
141 patterns of protein evolution. In addition, to help assess underlying causes of detected
142 patterns, we also evaluated the influence of the magnitude and breadth of gene expression on
143 observed rates of protein evolution. Our goals were to understand the forces shaping the
144 molecular evolution of reproductive proteins in this plant group, and whether this differs from
145 evolutionary patterns in other protein types and/or those proposed to shape reproductive
146 protein evolution in animal systems.
147
148
149 **2 Materials and Methods**
150
151 **2.1 Study Species**
152
153 Our analyses used gene expression and sequence data from each of four species within the wild
154 tomato group (*Solanum* section *Lycopersicum*): *S. lycopersicum* (domesticated tomato; LA3475)
155 and three wild relatives—*S. pimpinellifolium* (LA1589), *S. pennellii* (LA0716), and *S. habrochaites*
156 (LA1777) (Figure S1). (LA#### refers to the specific germplasm accession ID, as described in
157 tgrc.ucdavis.edu, from which all or most of the data were obtained; Table S1.) This group is a
158 clade of 12 closely-related diploid wild species along with the domesticated tomato, within the
159 hyperdiverse (~1300 species) plant genus *Solanum* (Sarkinen et al. 2013). All 12 wild tomato
160 species arose within the last ~2.5 million years, consistent with recent rapid speciation in this
161 clade (Pease et al. 2016a).
162
163 Mating system varies substantially across species in *Solanum* (Goldberg et al. 2010) and our
164 four focal species differ in their mating system, most notably in whether species are historically
165 self-incompatible (SI: *S. pennellii* and *S. habrochaites*) versus self-compatible (SC: *S.*
166 *lycopersicum* and *S. pimpinellifolium*). These species also have large estimated differences in
167 their genetic effective population sizes (Ne), that are consistent with these SI vs SC mating
168 system differences. For example, prior estimates of average heterozygosity from transcriptome-
169 wide data are 0.28% and 0.245% for *S. pennellii* and *S. habrochaites* accessions respectively,
170 and 0.04% for *S. pimpinellifolium* accessions (Table S3 in Pease et al. 2016a). Heterozygosity is
171 even lower in the domesticated tomato (*S. lycopersicum*) (e.g., Rick and Fobes 1975). In
172 *Solanum*, genetic self-incompatibility (the inferred ancestral state) is based on molecular
173 interactions between proteins expressed in growing pollen tubes and pistil-expressed SRNase
174 and other known proteins (Cruz-Garcia et al. 2003; Takayama and Isogai 2005). SI is observed as
175 arrest of pollen tube growth in the female reproductive tract (the pistil). Self-compatible
176 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 pollen production on average, in comparison to SC species (Vosters et al. 2014), even though the transition to SC likely occurred within the last ~0.5MY (the estimated age of the clade containing the two SC species here; Pease et al. 2016a). Both differences in the capacity to reject self pollen (i.e. SI versus SC), and these morphological differences, could have large effects on the average diversity of mating partners, and therefore the opportunity for and magnitude of intraspecific sexual selection and sexual conflict (Willson 1994, Delph and Havens 1998, Clark et al. 2006). In particular, while cross-pollination rates in SI species must be 100%, these have been estimated to range from 0-40% in wild SC species, including *S. pimpinellifolium* (e.g., Rick et al. 1977, 1978).

2.2 Quantifying gene expression and determining tissue-specificity

Our analyses used gene expression (RNAseq) data collected individually from leaf (up to five developmental stages), root, seed, vegetative meristem, stem, flower (several stages), style, pollen, and ovule tissues in each accession. RNAseq data were obtained from seven publically available SRA projects (Table S1), mostly drawn from three previously published analyses (Koenig et al. 2013, Ichihashi et al. 2014, Pease et al. 2016b), except data from ovule tissue which has not been previously published. Note that style and pollen data were not available specifically for *S. pimpinellifolium*. Sources of each dataset are described in detail in Table S1. Procedures for generating the ovule RNAseq data are described in the supplementary text. Each tissue had 1-6 replicates (generally 3) per accession (Table S1).

For each library within each tissue, we processed the raw reads by filtering adapter sequences and low quality bases with TRIMMOMATIC (Bolger et al. 2014). Reads were then mapped against the tomato reference genome (ITAG 2.4) using STAR (Dobin et al. 2013), with default settings excluding alignments with >10 mismatches. Numbers of reads mapped onto genic regions were estimated with FEATURECOUNTS (Liao et al. 2014). We normalized the read counts from each library by calculating TPM (transcripts per million) and then calculated the mean normalized read counts for each tissue of each species.

We first classified loci into three general classes of gene—reproductive (RP), vegetative (VG), and general (GR) (i.e., expressed in both reproductive and non-reproductive/vegetative tissue types). Reproductive genes were required to be expressed (TPM>2) in at least one reproductive tissue (i.e. style, pollen, ovule), and to have no or trace expression (TPM <0.5) in any of the remaining tissues (i.e. leaf, root, stem, seed, root, vegetative meristem), in at least two of the three investigated species for which we had all tissues available (i.e. *S. lycopersicum*, *S. pennellii*, and *S. habrochaites*). (Because *S. pimpinellifolium* lacked some of the tissue types in our RNA-seq data, we did not include this species in the filtering/determination of tissue-specificity.) Vegetative genes were defined similarly by requiring expression in at least one

221 vegetative (non-reproductive) tissue, and no/trace expression in any reproductive tissue.
222 Generally-expressed (GR) genes were expressed in at least one reproductive tissue and also in
223 at least one vegetative tissue.
224
225 Using similar criteria, we also identified a set of tissue-specific genes, defined as those only
226 expressed in one specific tissue (TPM>2) and not in any other tissue (TPM<0.5) in at least two
227 out of the three species for which we had all tissue types (i.e. *S. lycopersicum*, *S. pennellii*, and
228 *S. habrochaites*). For these analyses, we focused on specific tissues (i.e. leaf, root, stem, seed,
229 vegetative meristem, style, pollen, ovule) for which we had sufficient sampling for at least three
230 species (Table S1). Finally, using the same criteria and the expression domains in the three
231 species for which we had all tissues represented, we also classified transcripts as gametophytic-
232 exclusive (pollen and/or ovule expression only) versus sporophytic-exclusive versus both
233 gametophytic and sporophytic (pollen and/or ovule expression, and at least one other diploid
234 tissue), as well as ovule-specific versus expressed in both ovules and at least one sporophytic
235 (diploid) tissue, and pollen-specific versus expressed in both pollen and at least one sporophytic
236 (diploid) tissue.
237
238 **2.3 Estimating rates of protein evolution**
239
240 As with prior studies (Clark et al. 2006 and references therein), we used dN/dS—the ratio of the
241 number of non-synonymous substitutions per non-synonymous site to the number of
242 synonymous substitutions per synonymous site—as our estimate of protein evolution. Although
243 dN/dS is usually <1—because most functional loci experience some constraint on non-
244 synonymous changes—larger values of dN/dS are consistent with faster protein evolution. Note
245 that estimates of dN/dS can be influenced by intraspecific polymorphism when it is high
246 compared to fixed differences between species (see Hahn 2018, Chapter 7). However,
247 estimates of heterozygosity within each of our wild accessions are small compared to
248 divergence (e.g. dS) estimates across these species (Supplementary text), indicating that most
249 of the SNPs contributing to our estimates of dS, dN, and dN/dS, are fixed differences.
250
251 We estimated dN/dS for each locus in our dataset, to calculate mean rates of protein evolution
252 in different classes of genes, and to identify the set of loci with estimated dN/dS>1—a pattern
253 consistent with positive selection for protein evolution at that locus. In addition, to compare
254 the overall dN/dS for loci on SI versus SC branches, we compared estimates from sequences
255 that were concatenated across groups of genes (described below). Note that low transcript
256 coverage reduces the power to accurately call SNPs at heterozygous sites and can thereby
257 introduce errors into inferred sequences and statistical estimates based on these; this can be a
258 concern in low coverage RNA-seq data. However, the average per locus raw read count in our
259 dataset ranges from ~160 to >800 (Table S2), indicating that we have ample power to
260 accurately call SNPs with these transcript data.
261
262 To generate the sequence alignments for these analyses, we combined the RNA-seq data used
263 for quantifying gene expression (above) with whole-transcriptome data from the same four
264 accessions previously generated in Pease et al. (2016a), and with publicly available DNA-seq

265 data (Table S3); the latter datasets allowed us to augment our RNAseq data with additional
266 sequences for loci whose expression was not detected for one or more of our four species.
267 These reads were aligned against the tomato reference genome by STAR, with defaults
268 excluding alignments with >10 mismatches, as with the previous reference-aligned single copy
269 ortholog dataset (Pease et al. 2016a) The aligned data were processed using SAMTOOLS
270 followed by MVFTOOLS to generate the sequence alignments of orthologous genes. We only
271 retained transcript sequence alignments longer than 200 bps after removing gaps and indels,
272 leaving us with 21,216 sequence alignments (loci) for downstream analyses.
273
274 For each locus, we used PAML (Yang 2007) to estimate dN/dS across the branches of a given
275 phylogenetic topology among the four investigated species, using the input species tree
276 ((*lycopersicum*, *pimpinellifolium*), (*pennellii*, *habrochaites*)) which reflects the known species
277 topology (Figure S1). To do so we used the one-ratio (M0) codeml model in PAML, that
278 generates a single dN/dS estimate across all branches for each locus (that is, dN/dS is fixed at
279 one ratio for each gene). Outputs from this model were used to compare the mean and
280 distribution of dN/dS for genes among three broad categories (reproductive (RP), vegetative
281 (VE), and general (GR)), as well as between each of six tissue-specific groups classified by their
282 expression in only a single tissue (pollen, style, ovule, leaf, root, or stem; vegetative meristem
283 and seed were excluded from the latter analyses because they had no exclusively tissue-specific
284 genes), and between loci that had gametophytic versus sporophytic, or
285 gametophytic+sporophytic, domains of gene expression. We also compared loci expressed in
286 ovules only versus expressed in ovules and at least one sporophytic (diploid) tissue, or in pollen
287 only versus expressed in pollen and at least one sporophytic tissue.
288
289 Second, in addition to locus-by-locus dN/dS comparisons, we also evaluated overall dN/dS
290 estimates along SC branches in comparison to those along SI branches. Among our four species,
291 the two self-compatible species (*S. lycopersicum* and *S. pimpinellifolium*) are most closely
292 related to each other, and the lengths of SC branches (those leading to *S. lycopersicum* and *S.*
293 *pimpinellifolium*) are much shorter than the length of SI branches (those leading to *S. pennellii*
294 and *S. habrochaites*) (Figure S1). As a result, branch lengths (especially for SC lineages) were too
295 short to reliably estimate lineage-specific dN/dS for each locus individually. Therefore, for these
296 analyses we concatenated all loci within each class of genes to be compared, and then
297 estimated dN/dS across the concatenated set of loci for SI and SC branches separately. For
298 example, for reproductive (RP) loci, we concatenated all RP loci and estimated dN/dS across the
299 entire concatenated set, allowing for different dN/dS estimates (variable rates) on SI versus SC
300 branches. Because of the small numbers of loci concatenated for some of our individual tissues,
301 we limited SI vs SC branch-specific estimates to our three broad classes of loci (RP, VG, and GR)
302 only.
303
304 Because the phylogeny of our species contains one internal branch, for these analyses we used
305 two alternative multi-rate (M2) codeml models in PAML. For the first, three-ratio, model we
306 allowed the internal branch to have a rate that differed from the tip branches leading either to
307 SC species, or to SI species. For the second, two-ratio, model, we classified the internal branch
308 as 'SI' (i.e. dN/dS was estimated separately for tip branches leading to SC species, versus a

309 background rate for the tip branches leading to SI species plus the internal branch). This
310 approach was used to estimate branch-specific rates (dN/dS estimates) for SI versus SC
311 branches for each of our three general classes of loci (RP, VG, and GR). Note we report three-
312 ratio results in the main text, and two-ratio results in the supplement.
313
314 **2.4 Comparing protein evolution rates**
315
316 All statistical analyses were performed in R version 3.6.3 (R Core Team 2015). For all
317 comparisons, we first removed genes showing unusually high values (dN/dS >10) from the PAML
318 output, as these were due to estimation limitations within PAML (e.g. a lack of synonymous
319 substitutions across the branch(es) used for estimation) or otherwise from poor sequence
320 quality or alignment.
321
322 To compare the mean rate of per locus protein evolution (dN/dS) between different classes of
323 loci, we fit generalized linear models (GLMs). Shapiro-Wilk tests of normality and quantile-
324 quantile plots were used to assess normality in the distribution of dN/dS; in all categories,
325 dN/dS was heavily skewed towards lower values and zero-inflated (reflecting the fact that most
326 expressed genes are under purifying selection). To accommodate this in our models, we assume
327 a gamma residual distribution with an identity link function, rather than the standard Gaussian
328 assumption, after confirming the suitability of this distribution with maximum likelihood (using
329 the fitdistrplus R package). Following GLMs, to compare the mean estimated dN/dS between
330 specific classes of loci, we made pairwise comparisons using Tukey post-hoc tests, adjusting p-
331 values for multiple comparisons. Specifically, post-hoc test p-values were adjusted using the
332 default 'single-step' method in the R package 'multcomp' (Hothorn et al. 2008), which adjusts
333 p-values based on the joint normal or t distribution of the linear function. In addition, we
334 compared the proportion of positively selected genes (i.e. dN/dS>1) between different classes
335 of loci by performing Chi-Squared tests of independence and, for specific tests of enrichment in
336 single categories (e.g., in RP relative to GR), one-sided Fisher's exact tests.
337
338 To assess evidence for differences in dN/dS between mating systems, we aimed to compare
339 dN/dS estimates from concatenated data, as calculated separately for SI and SC branches. Both
340 biological and technical factors make these comparisons challenging to interpret. Biologically,
341 overall differences in Ne and therefore in the expected efficacy of positive and purifying
342 selection, could generate global differences in rates of protein evolution between SC and SI
343 lineages (see Introduction), that are unrelated to their potential differences in the historical
344 strength of sexual selection. Technically, in our specific dataset, the two SC species are much
345 more closely related than the two SI species, so there are large differences in branch lengths
346 (number of substitutions) over which we are estimating dN/dS values for SI versus SC branches.
347 Therefore, to assess whether exclusively reproductive (RP) loci had a different pattern of
348 molecular evolution on SI versus SC branches in comparison to other loci, we compared the
349 observed difference in dN/dS between SI and SC branches for RP genes to an estimate of the
350 'baseline' genome-wide difference in dN/dS between SI and SC branches. This baseline was
351 described by the SI vs SC difference in our set of generally expressed genes (GR), which is
352 expected to be due to factors unrelated to variation in sexual selection. Because we have only a

353 single estimate (from the concatenated data) for each class of gene for each branch type,
354 standard parametric tests can't be used. Instead we used bootstrap resampling to estimate the
355 variance around the mean SI vs SC difference in dN/dS in the GR dataset, and then to evaluate
356 whether the observed SI vs SC difference in RP genes falls outside the range of this variance (i.e.
357 is larger or smaller than the 95% CI of this distribution). For each bootstrap replicate (N=1000
358 replicates), we re-sampled 500 loci at random from the GR dataset, estimated dN/dS for SI and
359 SC branches and the difference between these two estimates. Our observed dN/dS difference
360 (SC-SI rate) was compared to the resulting distribution of 1000 replicate dN/dS differences from
361 GR loci.
362
363 Finally, to confirm that detected differences in dN/dS values between classes of genes were not
364 driven by differences in rates of synonymous mutation—which, for example, can systematically
365 vary according to genomic location or nucleotide content—we also evaluated differences in dN
366 and dS separately for these classes of genes (Supplementary text).
367
368 **2.5 Assessing the effect of gene expression level on molecular evolutionary rates**
369
370 Numerous studies have shown that rates of protein evolution can be influenced by the
371 magnitude and breadth of gene expression, such that highly expressed genes and/or genes
372 expressed in a broader number of tissues tend to have systematically lower estimated dN/dS
373 (Meisel 2011). To assess the effect of the magnitude and breadth of gene expression on our
374 inferences, we also compared average dN/dS values between genes expressed in different
375 numbers of tissues (ranging from 1 (tissue-specific loci) through to 8 (expressed in all tissues
376 examined)), using the criteria that a gene be expressed >2 TPM in a specific tissue in all species
377 for which we had data on that tissue-type. Within each of the broad classes of RP, VG, and GR
378 loci, we also assessed the quantitative relationship between mean gene expression level
379 (averaged across all tissues in which a locus was expressed) and estimated dN/dS. For this
380 analysis we constructed a multiple regression model of dN/dS (square root transformed to
381 approximate normality), including both mean gene expression (TPM), tissue class (RP, VG, or
382 GR), and their interaction term as independent variables.
383
384 **2.6 Assessing enrichment of functional categories among adaptively evolving loci**
385
386 To test for functional enrichment among our loci with dN/dS > 1, we performed a Gene
387 Ontology (GO) term enrichment analysis using Panther version 14 (Mi et al., 2019; Ashburner et
388 al., 2000; The Gene Ontology Consortium, 2019). We performed separate tests for GR, VE, and
389 RP tissue categories, as these classes had large sample sizes appropriate for tests of
390 enrichment. Loci were evaluated using both the biological process and molecular function
391 annotation data sets, and significance was determined using Fisher's Exact tests with false
392 discovery rate correction. False discovery rates were calculated using the Benjamini-Hochberg
393 procedure, using a critical value of 0.05 to filter results. All statistics were calculated using the
394 automated PANTHER Overrepresentation Tests.
395
396

397 **3 Results**
398
399 **3.1 Reproductive genes evolve at modestly faster rates, and have proportionally more**
400 **genes under positive selection**
401
402 Reproductive (RP) genes had an elevated mean rate of protein evolution compared to loci
403 expressed in both reproductive and non-reproductive (general/GR) tissues (Posthoc Z=2.548, p
404 = 0.0108) (Figure 1; Table 1; Table S4, Table S5). The mean dN/dS for vegetative-specific (VG)
405 loci fell between these two other classes of genes and did not differ from either of them (Table
406 1). These observed differences were not driven by variation in synonymous mutation rates (dS)
407 between these classes of genes (Supplementary text S3; Figure S2, Table S6). In addition, the
408 proportion of genes with dN/dS>1 differed by group (Table 1). A significantly greater proportion
409 of reproductive loci (28 of 670, or ~4.17%) are inferred to be evolving adaptively compared to
410 general loci (2.45%) (Fisher's exact test, p=0.0067; Table 1, Table S5). The proportion of
411 vegetative loci evolving adaptively (2.81%) was marginally lower than reproductive loci (Fisher's
412 exact test, p=0.093), and did not differ from general loci (Table 1; Table 4).
413
414 **3.2 Female- rather than male-specific genes evolve the most rapidly**
415
416 Comparisons among tissue-specific genes indicated no evidence that pollen-specific loci evolve
417 more rapidly than loci from female-derived tissues (style and ovule) (Figure 2; Table 2: Table
418 S7). Instead, style- and ovule-specific loci had the highest mean dN/dS values of all tissue
419 groups, and posthoc pairwise tests indicated that style mean rate was significantly greater than
420 rates in root-specific and leaf-specific loci, and ovule mean rate was significantly greater than
421 leaf mean rate (Figure 2; Table 2; Table S8). In comparison, pollen loci had a mean rate that was
422 intermediate between these groups of tissue-specific genes, and statistically indistinguishable
423 from each other tissue (Table S8). Note that many fewer loci were available for tissue-specific
424 comparisons, especially leaf- and stem-exclusive loci (Table 2), and two tissues—vegetative
425 meristem and seed—had no detected tissue-exclusive loci.
426
427 Similar to comparisons of mean rates, the proportion of genes with dN/dS>1 differed by tissue
428 (Table 2). Moreover, this proportion was significantly higher in style tissue—where 6 of 25 loci
429 were inferred to be evolving adaptively—compared to all other tissues (Fisher's exact test, p=
430 5.046×10^{-5}; Table 2). For ovule-exclusive loci, the proportion adaptive evolving loci (3 of 71)
431 was also greater than in all other tissues (excluding style-specific loci) (Fisher's exact test, p=
432 0.0376) (Table 2).
433
434 **3.3 Gametophytic-exclusive proteins do not evolve more rapidly than sporophytic-**
435 **exclusive proteins, or proteins expressed in both tissue types**
436
437 Although the mean rate of protein evolution in gametophytic genes was slightly higher than the
438 rate of protein evolution in sporophytic-exclusive loci, this difference was not significant (t=-
439 0.452, P=0.588) (Table 3; Table S9). Gametophytic-exclusive proteins also do not evolve faster
440 than those expressed in at least one gametophytic and one sporophytic tissue (t=-1.498,

441 P=0.134) (Table 3). Similar to mean rate comparisons, the proportion of gametophytic loci with
442 dN/dS>1 (3 of 86, or 3.49%) did not differ from this proportion in sporophytic-exclusive genes
443 (68 of 2033, or 3.34%), and was slightly but not statistically larger than this proportion in
444 gametophytic+sporophytic genes (383 of 15778, or 2.43%) (Table 3).
445
446 As observed in the general gametophytic comparisons, mean rate of protein evolution in ovule-
447 exclusive loci was slightly higher than the rate of protein evolution in loci expressed in ovules
448 and at least one sporophytic tissue (Table 3), but this difference was not significant (GLM
449 t=1.212, p=0.225; Table S9). The proportion of ovule-exclusive loci with dN/dS>1 (3 of 71, or
450 4.23%) also did not differ from the proportion of ovule+sporophytic loci with dN/dS>1 (225 of
451 7884, or 2.85%) (Table 3). Neither the mean rate of protein evolution (GLM t=-0.037, p=0.971;
452 Table S9) nor the proportion of loci with dN/dS>1 (Table 3) differed between pollen-exclusive
453 loci and loci expressed in pollen and at least one sporophytic tissue.
454
455 **3.4 Patterns of protein evolution show mating-system effects that are not specific to**
456 **reproductive loci**
457
458 Consistent with genome-wide effects of lower Ne in our self-compatible lineages, rates of
459 protein evolution were inferred to be higher overall on self-compatible branches in both GR
460 and RP classes of loci, although not for VG loci (Table 4); in the latter case, VG rates were
461 numerically higher on SI branches. Results were similar for two-ratio models (Table S10). The
462 observed difference in rates between SC and SI branch for RP loci (0.0338) did not fall outside
463 the 95% CI of bootstrapped values (0.0007-0.0797) from the broader class of GR genes; instead
464 33% of bootstrapped samples had a dN/dS difference smaller than observed for RP loci.
465
466 **3.5 Higher rates of protein evolution are consistently associated with greater tissue-**
467 **specificity**
468
469 Across loci classified according to the number of tissues in which they were expressed at >2
470 TPM (from 1 to 8 tissues), mean dN/dS was highest in loci expressed in a single tissue and
471 consistently declined across loci expressed in increasingly more tissues (Figure 3; Table 5; Table
472 S11, S12). Similarly, both mean expression level and tissue class (GR, RP, VG) affected the per
473 locus estimates of protein evolution (dN/dS), as did their interaction (Table 6), indicating that
474 the relationship between expression level and rates of protein evolution varies depending upon
475 the broad tissue class. In particular, while the relationship between mean expression level and
476 dN/dS was significantly negative for generally-expressed (GR) loci, the 95% CI for this slope
477 overlapped zero for RP loci, indicating little evidence for a relationship between expression
478 level and rate of protein evolution for the RP class of loci (Figure S3; Table S13). Classes of
479 genes also differed in their overall mean gene expression (mean total TPM), which was highest
480 in RP genes (505.2±23.7) compared to GR genes (398.7 ±4.2), and VG loci (43.9±4.5) (Table
481 S14). Finally, mean gene expression level also differed between some tissue-specific loci, with
482 style and ovule loci having the highest average gene expression (Table S15).
483

These findings are generally consistent with greater evolutionary constraint on loci with broader expression domains, across all our loci. In addition, for genes expressed in at least one reproductive and non-reproductive tissue (i.e., general loci), they indicate greater constraint on genes with higher 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 S16).

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 elevated 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 reproductive protein evolution, and the factors proposed to drive these patterns.

4.1 Female reproductive proteins tend to evolve more rapidly

Our evidence for elevated rates of reproductive protein evolution was observed both as a modest but significant increase in mean dN/dS across all reproductive loci, and as a significantly larger proportion of loci showing evidence of adaptive divergence between species (dN/dS>1). Interestingly, however, this elevated evolution appears to be more clearly associated with proteins that have female- rather than male-specific functions. Indeed, female reproductive loci predominate among our most rapidly evolving loci, both generally and in analyses of tissue-specific loci. Of the tissue-specific loci found to have dN/dS>1, all are expressed in a female-specific tissue (either style or ovule) (Table S16).

527 This finding is intriguing. Some previous studies have detected accelerated evolution in proteins
528 with specific female-associated functions, for example, egg proteins involved in sperm-egg
529 interactions (e.g. Galindo et al. 2003). Nonetheless, most studies that include both male and
530 female proteins have found mixed evidence for elevated female-biased or female-specific
531 evolution (e.g. Wong 2010, Haerty et al. 2007; including in plants, e.g. Gossman et al. 2016),
532 compared to much clearer accelerated evolution in male-specific or male-biased proteins
533 (reviewed in Wong 2011, Wilburn and Swanson 2016). Some of this male-female difference
534 could simply reflect a comparative lack of knowledge about female-associated proteins, leading
535 them to be systematically underexamined in many comparisons. Indeed, most prior studies in
536 animals (reviewed, for example, in Clark et al. 2006, Dapper and Wade 2020), as well as some in
537 plants (Arunkumar et al. 2013, Harrison et al. 2019, and references therein), have focused
538 largely or entirely on reproductive proteins with male-specific functions. Here, our loci were
539 defined by tissue-specific patterns of gene expression, rather than from *a priori* expectations of
540 function, which might have helped to ameliorate any such biases.
541
542 Our data also indicates a role for adaptive evolution in the observed acceleration of female
543 protein divergence. Elevated rates of non-synonymous change (but values of dN/dS still less
544 than 1) could result from relaxed purifying selection (see further below) or because positive
545 selection has only acted on a subset of sites within each gene, or a combination of these
546 effects. However, we also found a larger proportion of female (style and ovule) loci with
547 dN/dS>1, suggesting a concentration of adaptive female protein evolution specifically involving
548 post-pollination reproductive processes. One driver of this could be sexual interactions. In many
549 animals, where female choice can be exerted both before and after mating (the latter as
550 'cryptic female choice') (Andersson and Simmons 2006), however in angiosperms male-female
551 (pollen-pistil) interactions can only act after pollen arrives at a flower. Therefore, all
552 evolutionary dynamics involving female choice or male-female interactions must act upon loci
553 expressed in this relatively narrow post-mating window. Indeed, among our rapidly evolving
554 (dN/dS>1) style loci, we detect at least one protein with putative functions in pollen
555 interactions (a pistil-specific extension-like protein; Solyc02g078080), in addition to several
556 biosynthetic enzymes (Table S16). (In contrast, 2/3 rapid ovule-specific genes have potential
557 functions in stress response, possibly suggesting roles that are not directly related to sexual
558 interactions—see below.) Moreover, in other angiosperms, 'female' loci that show rapid
559 adaptive divergence are also mostly pistil-specific proteins involved in selection among
560 different pollen genotypes (via roles in genetic self-incompatibility) (Clark et al. 2006), again
561 suggesting the importance of post-pollination sexual dynamics in accelerated reproductive
562 protein evolution.
563
564 Whether this potential adaptive explanation extends to explain the general elevation of female
565 reproductive protein evolution (Figure 2) or reproductive proteins overall (Figure 1) is not yet
566 clear. One prior study that compared protein evolution in male, female, and non-reproductive
567 tissues in *Arabidopsis thaliana* and two relatives, also found that mean dN/dS was elevated in
568 loci with expression biased towards female tissues (egg cell, central cell, and synergid cells:
569 three haploid tissues within the female gametophyte) but not in pollen-biased proteins
570 (Gossman et al. 2014). This pattern was attributed to relaxed selection on female-biased loci,

571 partly because additional site-specific tests did not consistently detect evidence of positive
572 selection in these loci (Gossman et al. 2014). Because our four species are closely related, we
573 do not have sufficient power to assess site-specific evidence for adaptive evolution in loci
574 whose estimated dN/dS is less than 1. Future analyses that pair divergence-based estimates
575 with polymorphism data within species—that can provide alternative and often more powerful
576 tests of positive and relaxed selection (e.g. Gossman et al. 2016, Arunkumar et al. 2013)—will
577 be helpful in differentiating the relative influence of these different forms of selection on the
578 broader patterns of elevated dN/dS detected here.
579
580 Regardless, if elevated female protein evolution is at least partly driven by post-pollination
581 sexual interactions, it is interesting that we do not also observe elevated evolution in pollen-
582 acting proteins. Of prior analyses in angiosperms, several have detected broadly elevated
583 evolution in proteins that are pollen specific (e.g., Arunkumar et al. 2013, Gossman et al. 2016,
584 Harrison et al. 2019). In principle, pollen protein evolution could be more constrained than style
585 or ovule evolution due to factors like differences in the degree of tissue-specificity or in the
586 operation of haploid selection, but these explanations are not generally consistent with our
587 data (see further below). Alternatively, it's possible that our analysis did not capture all the
588 male proteins relevant to post-pollination sexual interactions. For instance, pollen loci whose
589 expression requires direct interactions with the style and/or other pollen tubes would not be
590 captured in our expression atlas (which was generated from RNA expression in mature
591 ungerminated and *in vitro* germinated pollen; Table S1). There is evidence in other angiosperms
592 that pollen transcriptomes can be actively modified during their transit through the style,
593 resulting in changes in the suite of proteins being expressed (e.g., Qin et al. 2009, Mizukami et
594 al. 2019). Accounting for and including these genes would require transcriptomes from 'mixed'
595 post-pollination tissue, as well as specific approaches to differentiate pollen- from stylar-
596 expression in these mixed samples (e.g. Pease et al. 2016b). Importantly, this limitation would
597 also extend to any style- or ovule-specific proteins that are only elicited during pollination, so
598 may not be a complete explanation for the general difference we detect between male and
599 female loci here.
600
601 Other explanations for the under-accounting of pollen loci could include technical effects such
602 as detection limits on gene expression in pollen, and/or exclusion of pollen-expressed loci that
603 are especially short or rapidly evolving, due to filtering during transcript-mapping steps. In
604 addition, by using dN/dS comparisons only, we would not have detected protein changes that
605 involve duplicate, chimeric, or novel genes, and/or if different loci are the targets of selection in
606 different lineages. If these latter factors do contribute to our failure to detect elevated pollen
607 protein evolution, it suggests that different dynamics shape pollen protein identity and/or
608 evolution in *Solanum* compared to other angiosperms that have shown evidence for elevated
609 evolution using metrics like dN/dS (e.g. Arunkumar et al. 2013, Gossman et al. 2016, Harrison
610 et al. 2019). Regardless, with the data that we currently have, we do not see evidence that male
611 functions are subject to especially intense sexual selection, particularly in comparison to female
612 reproductive proteins.
613
614 **4.2 Rates of protein evolution are affected by tissue-specificity but not haploid expression**

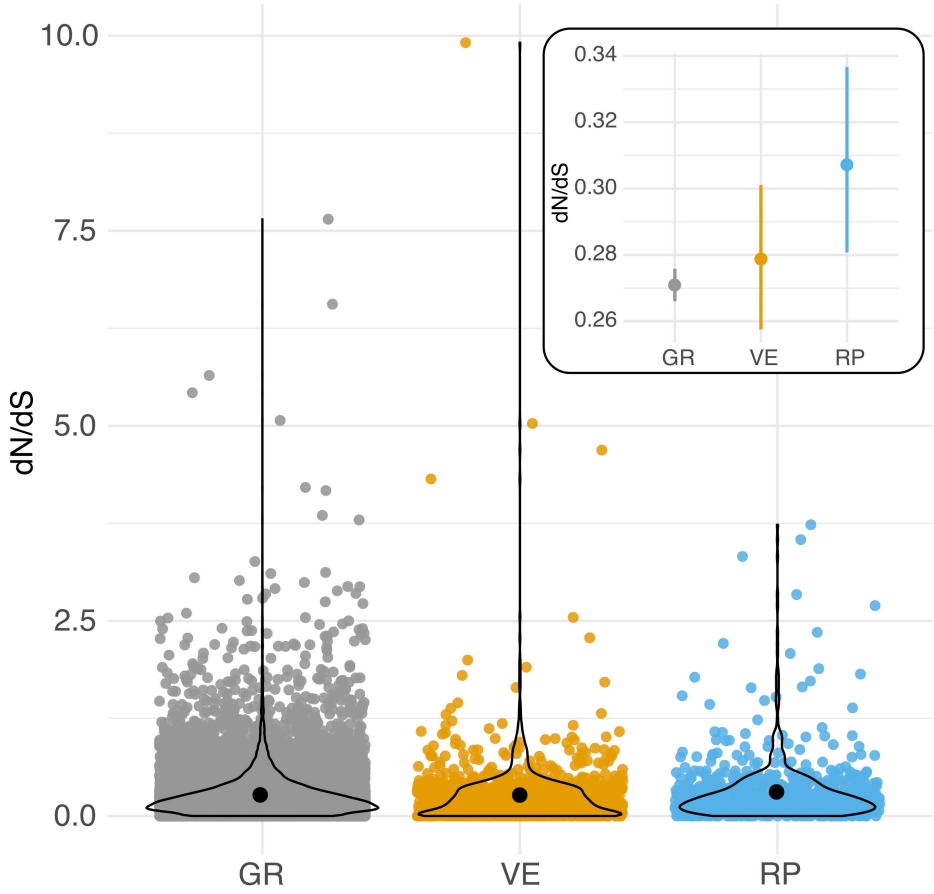
615
616 Our results can also suggest factors that shape the patterns of reproductive protein evolution
617 that we do observe, at least among the three broad factors (gene expression level, haploid
618 expression, and/or mating system variation) that we addressed here.
619
620 Of these, we find the strongest evidence that rates of protein evolution are influenced by the
621 breadth and magnitude of gene expression. Many studies have shown higher rates of protein
622 evolution in genes with more narrow expression domains and/or lower levels of gene
623 expression (Meisel 2011, and references therein), a relationship that is thought to reflect lower
624 constraint on tissue-specific proteins and stronger purifying selection on genes with high
625 expression levels or broad expression domains. We similarly found that higher rates of protein
626 evolution were associated with greater tissue-limitation and (in some cases) lower levels of
627 gene expression in our species. However, for reproductive protein evolution specifically, we
628 infer that tissue-limitation (a more narrow expression domain) is likely more important than
629 magnitude of gene expression. For the latter, gene expression was not significantly associated
630 with protein evolution for reproductive genes (Figure S3, Table S13), and reproductive loci also
631 had much higher mean gene expression levels in comparison to generally-expressed genes and,
632 especially, vegetative loci (Table S14).
633
634 High tissue-specificity has been proposed as an important contributor to elevated rates of
635 reproductive protein evolution—and a better predictor of this than sex-biased gene expression
636 *per se*—in both animals (Meisel 2011) and some plants (Veltsos 2019). For example, in
637 *Arabidopsis thaliana* and close relatives, Harrison et al. (2019) found faster protein evolution in
638 both pollen-specific and tissue-specific sporophytic genes compared to loci expressed in >1
639 tissue, although pollen-specific loci still retain the highest evolutionary rates in this and other
640 analyses in these species (e.g. Szovenyi et al. 2013, Gossman et al. 2016). Arunkumar et al.
641 (2013) also inferred a role for tissue-limitation or specificity in elevated dN/dS in *Capsella*
642 *grandiflora* pollen proteins (although this elevation was attributed to different forms of
643 selection—relaxed purifying selection in pollen germ cell genes versus greater purifying and
644 positive selection in pollen tube loci—depending on the specific sub-tissues within pollen).
645 These prior inferences, and our observations, suggest that higher tissue-specificity could
646 facilitate broadly elevated rates of protein evolution. Nonetheless, tissue-specificity cannot
647 provide a complete explanation of elevated reproductive protein evolution in our dataset, as
648 tissue-specific non-reproductive proteins still evolve more slowly than tissue-specific (female)
649 reproductive proteins (Table 2).
650
651 In comparison, we found no rate differences between loci expressed in haploid versus diploid
652 tissues, and therefore little evidence for a strong effect of haploid selection in driving
653 differences between reproductive and non-reproductive proteins. Haploid selection has two
654 projected effects on haploid-expressed loci—increasing the efficacy of positive selection on
655 advantageous alleles, and of purifying selection on deleterious alleles (Immler and Otto 2018,
656 Mattila et al. 2018)—compared to diploid-expressed loci. We have limited ability to
657 differentiate these two effects with our particular estimate of protein evolution, but our
658 observations are not consistent with a strong effect of either. In terms of accelerating adaptive

659 protein evolution, we found no significant elevation of evolutionary rates in loci limited to
660 gametophytic-tissues in general, and in ovule-limited and pollen-limited loci specifically (Table
661 3). Similarly, we find little evidence for stronger or more widespread purifying selection on
662 haploid-expressed loci, which would be observed as lower mean rates of protein evolution in
663 gametophytic loci. Stronger purifying selection due to haploid expression has been proposed as
664 one reason why sex-biased loci (generally, pollen expressed loci) do not appear to evolve faster
665 than non-sex-biased loci in several dioecious species (Sanderson et al. 2019, Cossard et al.
666 2019; discussed in Muyle 2019, but see also Veltsos 2019) and, indeed, sometimes appear to
667 evolve more slowly (e.g., Darolti et al. 2018). Our data do not suggest stronger purifying
668 selection in either gametophytic-limited proteins in general, or ovule- and pollen-limited loci
669 specifically, in our species. However, because we have relatively few pollen-specific loci
670 (discussed above) we cannot exclude the possibility that intermediate rates of pollen protein
671 evolution result from a tension between both positive and purifying selection predicted to act
672 on these loci. Nonetheless, our present data indicates little evidence for a global effect of
673 haploid gene expression on molecular evolutionary rates.
674
675 **4.3 Mating system variation does not does not differentially influence reproductive**
676 **protein evolution**
677
678 Our results also did not support a strong consistent effect of mating system differences—and
679 therefore predicted differences in the strength of sexual selection—on patterns of reproductive
680 protein evolution. Interpreting these comparisons is complex because differences in genetic
681 effective population size are expected to have global effects on rates of molecular evolution
682 (Charlesworth and Wright 2001, and see Introduction) that are unrelated to sexual selection *per*
683 *se*. The broad patterns of dN/dS we observed generally reflected these expected global effects
684 of differences in Ne due to mating system variation: in the SC lineages, mean dN/dS was higher
685 in both RP and GR genes, consistent with a general relaxation of selection in more selfing
686 lineages (Charlesworth and Wright 2001) regardless of whether or not loci were associated with
687 reproductive functions (Table 4). This genome-wide effect might have overwhelmed more
688 subtle effects of mating system variation acting specifically on loci involved in male-male
689 competition and/or female choice. In addition, the transition to self-compatibility is relatively
690 recent (less than 1 MY) among our species, and self-compatible lineages still retain significant
691 potential for outcrossing (Rick 1979, Vosters et al. 2014), which generally ranges from 0-40%
692 (e.g. Rick et al. 1977) but has been estimated as high as 84% in some genotypes (Rick et al.
693 1978). These factors mean there may be only modest differences in the historical and current
694 opportunity for sexual selection among our lineages, and therefore limited opportunity to
695 observe systematic consequences of this transition as differential changes in evolutionary rates,
696 especially among a sample of four species. (Statistically, this is compounded in our particular
697 dataset by the two SC species branches being substantially shorter than those for our two SI
698 species.) This contrasts with some other angiosperms, such as *A. thaliana*, where the transition
699 to self-compatibility is about as old but the consequent effects on outcrossing rates has been
700 much more severe (outcrossing rates are frequently <5%; Harrison et al. 2019, and references
701 therein). Even among *Arabidopsis*, however, the effects of mating system on sexual locus
702 evolution might be complex. For example, Gossman et al. (2014) found that dN/dS in pollen-

703 acting genes was significantly lower in the outcrosser *A. lyrata* compared to selfing *A. thaliana*,
704 even though pollen competition could be expected to accelerate adaptive evolution in these
705 loci in the outbreeding species.
706
707 Interestingly, several previous analyses in animal systems have similarly found equivocal
708 evidence for differential effects of mating system variation on the evolution of reproductive
709 tract proteins (reviewed in Wong 2011). For example, in studies of primates, evolutionary rates
710 in specific sperm and copulatory plug proteins are positively associated with common proxies
711 for sexual selection, such as residual testis size and estimated number of mates (e.g. Dorus et
712 al. 2004, Martin-Coello et al. 2009). However other studies find that phylogeny-wide signals of
713 positive selection in reproductive tract proteins are not associated specifically with polyandrous
714 (as opposed to monogamous) lineages (e.g., Ramm et al. 2008, Finn and Civetta 2010; and see
715 Table 1 in Wong 2011). Limitations on phylogenetic models that can assess associations
716 between molecular evolutionary rates and changes in phenotypic characters, and variation
717 among lineages in the specific targets of sexual selection (Wong 2011), might explain relatively
718 weak effects of mating system variation on protein evolution rates, despite the importance of
719 sexual selection in driving faster reproductive protein evolution. Adding more species, and
720 focusing on pairs of SI-SC sister species, might help in addressing some of these limitations in
721 future analyses within *Solanum*.
722
723 Alternatively, the weak effect of mating system variation—both here and in prior studies—
724 might be because the predominant cause of globally elevated reproductive protein evolution is
725 not sexual selection. For example, pathogen resistance responses are also known to be
726 associated with elevated rates of protein evolution—driven by antagonistic host-pathogen
727 evolution—and might be concentrated among reproductive proteins if mating often leads to
728 the exchange of pathogens (Clark et al. 2006, Wong 2011). In principle, the identity of loci
729 inferred to be under positive selection could be helpful in differentiating among alternative
730 selective agents responsible for driving this adaptive divergence. Of the genes in our dataset
731 that show elevated evolution, we cannot yet draw strong conclusions about whether they
732 better reflect roles related to sexual interactions versus responses to non-sexual environmental
733 factors; many of them have unknown functions or general functions that do not *a priori* provide
734 specific support for either inference (Table S16). We do detect at least one style-specific protein
735 with a clear functional connection to pollen-pistil interactions. However, our rapidly-evolving
736 ovule-specific genes are implicated in stress responses, which might indicate a greater role for
737 natural rather than sexual selection in shaping adaptive responses in this class of loci.
738
739 **4.4 Consequences of rapid protein evolution for the evolution of reproductive differences**
740 **and speciation**
741
742 Finally, among the motivations for examining patterns of reproductive protein evolution is to
743 understand their possible role in lineage diversification. If reproductive proteins do indeed
744 evolve more rapidly than other classes of protein, then perhaps they play an outsized role in
745 the emergence of functional differences and reproductive isolating barriers between lineages.
746 For example, evidence suggests post-mating prezygotic traits frequently evolve rapidly in

747 animals (e.g. Galindo et al. 2003, Dorus et al. 2004, Plakke et al. 2019) and, when divergent
748 between species, could contribute to reproductive isolation (Howard 1999, Snook et al. 2009,
749 Palumbi 2009, Castillo and Moyle 2014). Similarly, divergence in post-pollination prezygotic
750 reproductive molecules could be important contributors to species barrier formation among
751 angiosperms (Moyle et al. 2014). Here we did detect evidence for a global elevation of protein
752 evolution in tissues that mediate these post-pollination prezygotic stages, especially of female-
753 specific loci. Our observations are consistent with a role for tissue specificity in facilitating these
754 elevated rates, but not for haploid selection or, as yet, for differences in selective dynamics that
755 are associated with recent shifts in mating system. The evolutionary factors that might drive
756 this more rapid evolution therefore remain to be clarified. Regardless, our results indicate that
757 rapid reproductive divergence is especially characteristic of proteins with female-specific
758 functions among our species. These specific functions can be critical for maintaining
759 coordinated sexual signals and ensuring fertilization post-pollination (Bernasconi et al. 2001,
760 Swanson et al. 2004, Moyle et al. 2014). Therefore their rapid divergence might be particularly
761 influential in the initial stages of reproductive isolation and, thereby, in speciation, in this highly
762 diverse plant genus.
763
764
765 **5 Author contributions**
766
767 LCM, MW, and MJSG designed the study. MW and MJSG performed the analyses. LCM wrote
768 the manuscript with assistance from MW and MJSG.
769
770
771 **6 Funding**
772
773 This work was supported by NSF grant IOS-1127059 and NSF grant DEB-1856469 to LCM.
774
775
776 **7 Acknowledgments**
777
778 Thanks to CJ Jewell for collecting ovule tissue, Christie Bergeron for RNA extractions, the
779 Indiana University Center for Genomics and Bioinformatics for assistance with library
780 construction and sequencing, and three anonymous reviewers for comments that helped
781 improve the study.
782

783
784



785
786
787 **Figure 1:** Per locus protein evolution (dN/dS) among loci expressed in reproductive (RP),
788 vegetative (VG), and both reproductive and non-reproductive (general/GR) tissues in *Solanum*
789 species. Black circles indicate means for each group. Inset: Estimated mean (circle) and 95% CI
790 of dN/dS for each group of loci. Overlaid violin plots show the distribution of each group of
791 data.
792

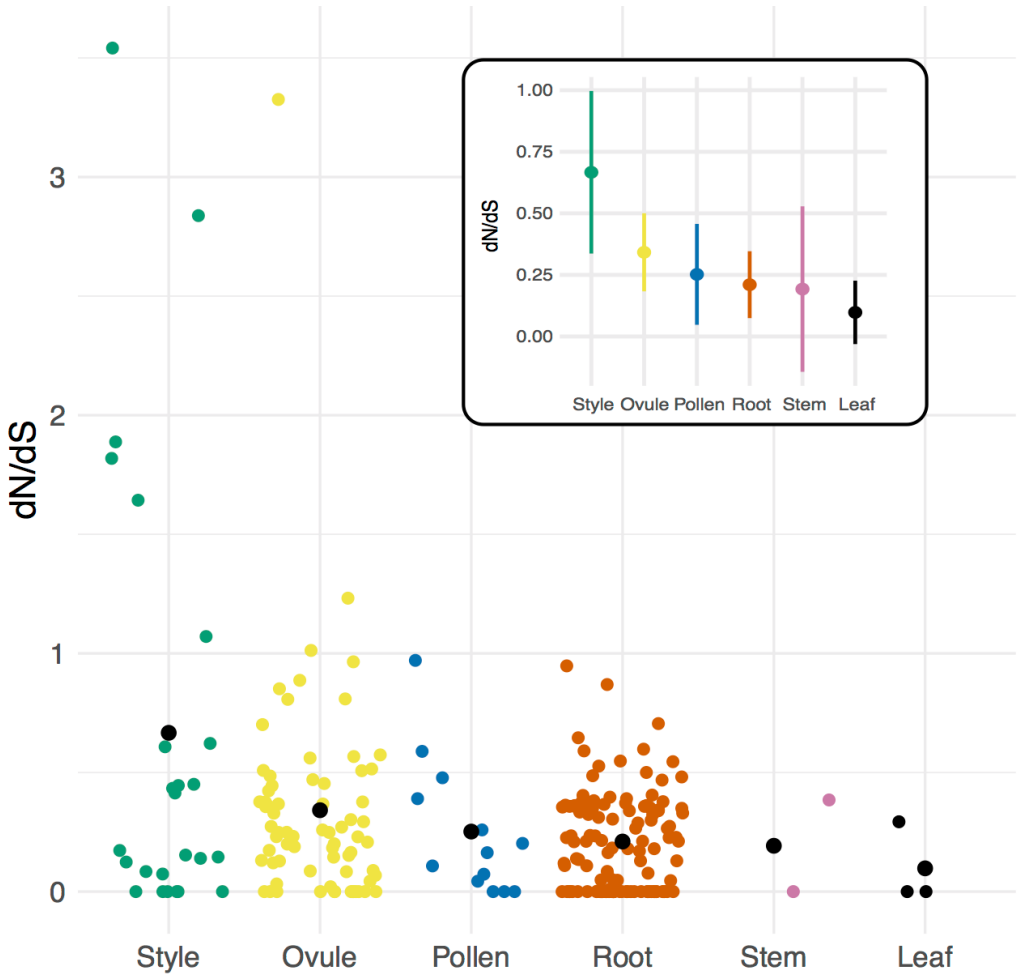


Figure 2: Per locus protein evolution (dN/dS) among groups of tissue-specific loci in *Solanum* species. Black circles indicate means for each group. Inset: Estimated mean (circle) and 95% CI of dN/dS for each group of loci.

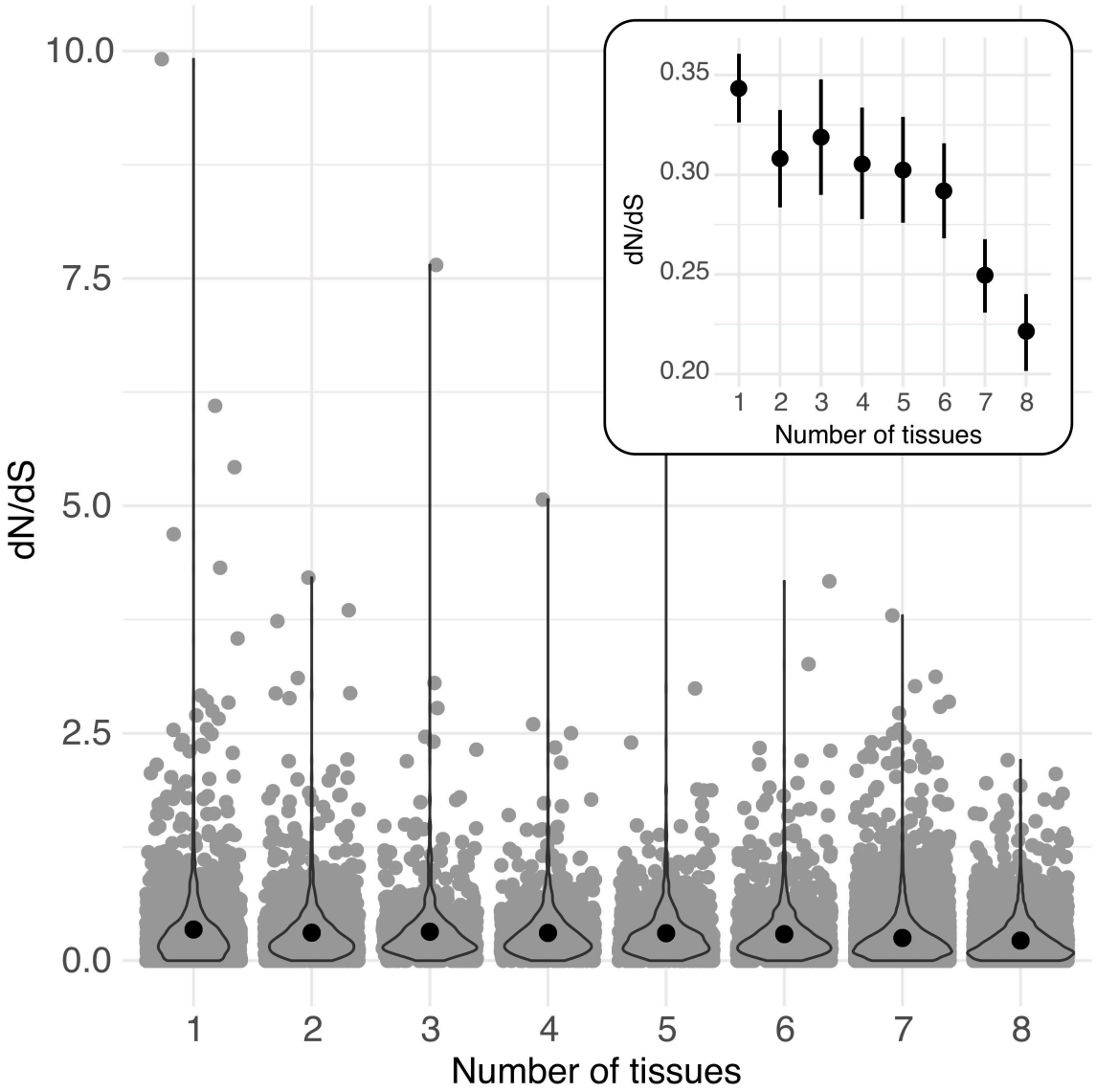


Figure 3: Per locus protein evolution (dN/dS) among groups of loci expressed (>2 TPM) in different numbers of tissues (from 1 to 8) in *Solanum* species. Black circles indicate means for each group. Inset: Estimated mean (circle) and 95% CI of dN/dS for each group of loci. Overlaid violin plots show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

808
809
810 **Table 1. Rates of protein evolution (dN/dS) in different broad classes of loci.** dN/dS calculated
811 for each locus using a one ratio test in PAML.
812

		Per locus dN/dS				Loci with dN/dS>1		
Class	N loci	mean	SE	Median	Tukey test[1]	Number	Proportion[2]	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

813
814 1. see Table S4.
815 2. The proportion of genes with dN/dS>1 differed among classes of loci (X^2=7.9696, p=0.0186).
816
817
818 **Table 2: Rates of protein evolution (dN/dS) in tissue-specific loci**
819

	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

820
821 1. see Table S8.
822 2. The proportion of genes with dN/dS>1 differed by tissue (X^2= 33.707, p=2.72 x 10-6). One-
823 sided Fisher's exact tests indicated this proportion was greater in style compared to all other
824 tissues (p=5.046 x 10-5), and greater in ovule compared to all other tissues except style
825 (p=0.03756).

826 **Table 3: Rates of protein evolution (dN/dS) in gametophytic (GM) versus sporophytic (SP) or**
827 **gametophytic+sporophytic (GM+SP) loci.** Within each GLM, tissue groups did not differ in
828 mean dN/dS (Table S9).
829

	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	

830
831 1. Fisher's exact tests
832
833
834 **Table 4: Estimates of dN/dS for SC and SI branches separately, for loci concatenated within**
835 **each broad class of genes.** Results are from 3 ratio model in PAML, with input species tree:
836 ((Slyc #1, Spim #1), (Spen #2, Shab #2)).
837

Class	# genes	# codons	Null model	Variable rate model	2*LnL	p value	SI-branch dN/dS	SC-branch dN/dS	Internal Branch	SC-SI diff.
GR	16221	6619178	-28576265	-28575900	729.61	<1 x 10x-16	0.2224	0.2617	0.1901	0.0392
RP	904	110171	-478937	-478932	11.02	0.00404	0.2285	0.2623	0.1999	0.0338
VG	1300	78355	-344551	-344546	9.31	0.00951	0.2267	0.2042	0.1850	-0.0225

838

839 **Table 5: Rates of protein evolution (dN/dS) in loci according to breadth of tissue expression.**
840

# 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

841
842 1. see Table S12
843
844
845 **Table 6: Relationship between mean gene expression level and estimated dN/dS, across loci**
846 **in each broad class (GR, RP, VG) of loci (Linear model)**
847

	DF	SS	MS	F	Pr > F
Expression level	1	20.53	20.5297	334.79	2.2 x 10-16
Class	2	2.69	1.3467	21.96	2.98 x 10-10
Expression level x Class	2	3.85	1.9239	31.37	2.50 x 10-14
Residuals	17808	1092.02	0.0613		

848
849

850 **8 References**

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

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

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

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

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

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

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

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

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

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

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

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

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

880 COSSARD GG, TOUPS MA, PANNELL JR. 2019 Sexual dimorphism and rapid turnover in gene
881 expression in pre-reproductive seedlings of a dioecious herb. Annals of Botany. 123(7):1119-
882 1131.

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

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

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

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

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

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

895 DORUS S, EVANS PD, WYCKOFF GJ, CHOI SS, LAHN BT. 2004 Rate of molecular evolution of the
896 seminal protein gene SEMG2 correlates with levels of female promiscuity. Nature Genetics.
897 36(12):1326-1329.

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

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

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

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

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

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

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

916 HAHN MW. 2018 *Molecular Population Genetics*. Sinauer/Oxford University Press, Oxford UK.

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

920 HOTHORN T, BRETZ F, WESTFALL P. 2008 Simultaneous inference in general parametric models.
921 Biometrical Journal. 50(3):346-363.

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

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

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

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

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

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

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

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

944 MARTIN-COELLO J, DOPAZO H, ARBIZA L, AUSIO J, ROLDAN ERS, GOMENDIO M. 2009 Sexual
945 selection drives weak positive selection in protamine genes and high promoter divergence,
946 enhancing sperm competitiveness. Proceedings of the Royal Society B-Biological Sciences.
947 276(1666):2427-2436.

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

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

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

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

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

961 MI H, HUANG X, MURUGANUJAN A, TANG H, MILLS C, KANG D, THOMAS PD. 2019 PANTHER

962 version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis
963 tools. . Nucleic Acids Research. 47(D1):D419-D426.

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

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

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

972 MOYLE LC, JEWELL CP, KOSTYUN JL. 2014 Fertile approaches to dissecting mechanisms of
973 premating and postmating prezygotic reproductive isolation. Current Opinion in Plant Biology.
974 18:16-23.

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

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

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

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

985 PLAKKE MS, WALKER JL, LOMBARDO JB, GOETZ BJ, PACEILA GN, DURRANT JD, CLARK NL,
986 MOREHOUSE NI. 2019 Characterization of female reproductive proteases in a butterfly from
987 functional and evolutionary perspectives. Physiological and Biochemical Zoology. 92(6):579-590.

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

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

993 RAMM SA, OLIVER PL, PONTING CP, STOCKLEY P, EMES RD. 2008 Sexual selection and the
994 adaptive evolution of mammalian ejaculate proteins. Molecular Biology and Evolution. 25(1):207-
995 219.

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

999 RICK CM, FOBES JF. 1975 Allozyme variation in cultivated tomato and closely related species

1000 Bulletin of the Torrey Botanical Club. 102(6):376-384.

1001 RICK CM, FOBES JF, HOLLE M. 1977 Genetic variation in *Lycopersicon pimpinellifolium* - evidence
1002 of evolutionary change in mating systems. . Plant Systematics and Evolution. 127(2-3):139-170.

1003 RICK CM, HOLLE M, THORP RW. 1978 Rates of cross-pollination in *Lycopersicon pimpinellifolium* -
1004 impact of genetic variation in floral characters Plant Systematics and Evolution. 129(1-2):31-44.

1005 RITCHIE MG. 2007 Sexual selection and speciation. Annual Review of Ecology, Evolution, and
1006 Systematics. 38:79-102.

1007 SANDERSON BJ, WANG L, TIFFIN P, WU ZQ, OLSON MS. 2019 Sex-biased gene expression in
1008 flowers, but not leaves, reveals secondary sexual dimorphism in Populus balsamifera. New
1009 Phytologist. 221(1):527-539.

1010 SARKINEN T, BOHS L, OLMSTEAD RG, KNAPP S. 2013 A phylogenetic framework for evolutionary
1011 study of the nightshades (Solanaceae): a dated 1000-tip tree. Bmc Evolutionary Biology. 13.

1012 SICARD A, LENHARD M. 2011 The selfing syndrome: a model for studying the genetic and
1013 evolutionary basis of morphological adaptation in plants. Annals of Botany. 107(9):1433-1443.

1014 SKOGSMYR I, LANKINEN A. 2002 Sexual selection: an evolutionary force in plants. Biological
1015 Reviews. 77(4):537-562.

1016 SLOTTE T, BATAILLON T, HANSEN TT, ST ONGE KR, WRIGHT SI, SCHIERUP MH. 2011 Genomic
1017 Determinants of Protein Evolution and Polymorphism in Arabidopsis. Genome Biology and
1018 Evolution. 3:1210-1219.

1019 SNOOK RR, CHAPMAN T, MOORE PJ, WEDELL N, CRUDGINGTON HS. 2009 Interactions between
1020 the sexes: new perspectives on sexual selection and reproductive isolation. Evolutionary Ecology.
1021 23(1):71-91.

1022 STEPHENSON AG, BERTIN RI. 1983 Male competition, female choice, and sexual selection in
1023 plants. In: *Pollination Biology*. (L R, ed), pp. 109-151. Academic Press, New York, NY.

1024 STONE JL. 2002 Molecular mechanisms underlying the breakdown of gametophytic self-
1025 incompatibility. Quarterly Review of Biology. 77(1):17-32.

1026 SWANSON R, EDLUND AF, PREUSS D. 2004 Species specificity in pollen-pistil interactions. Annual
1027 Review of Genetics. 38:793-818.

1028 SWANSON WJ, VACQUIER VD. 2002 The rapid evolution of reproductive proteins. Nature Reviews
1029 Genetics. 3(2):137-144.

1030 SWANSON WJ, VACQUIER VD. 2002 Reproductive protein evolution. Annual Review of Ecology
1031 and Systematics. 33:161-179.

1032 SZOVENYI P, RICCA M, HOCK Z, SHAW JA, SHIMIZU KK, WAGNER A. 2013 Selection Is No More
1033 Efficient in Haploid than in Diploid Life Stages of an Angiosperm and a Moss. Molecular Biology
1034 and Evolution. 30(8):1929-1939.

1035 TAKAYAMA S, ISOGAI A. 2005 Self-incompatibility in plants. Annual Review of Plant Biology.
1036 56:467-489.

1037 THE GENE ONTOLOGY CONSORTIUM. 2019 The Gene Ontology Resource: 20 years and still GOing
1038 strong. Nucleic Acids Research. 47(D1):D330-D338.

1039 TURNER LM, HOEKSTRA HE. 2008 Causes and consequences of the evolution of reproductive
1040 proteins. International Journal of Developmental Biology. 52(5-6):769-780.

1041 VACQUIER VD, SWANSON WJ. 2011 Selection in the Rapid Evolution of Gamete Recognition
1042 Proteins in Marine Invertebrates. Cold Spring Harbor Perspectives in Biology. 3(11).

1043 VELTSOS P. 2019 Not all sex-biased genes are the same. New Phytologist. 221(1):10-11.

1044 VOSTERS SL, JEWELL CP, SHERMAN NA, EINTERZ F, BLACKMAN BK, MOYLE LC. 2014 The timing of
1045 molecular and morphological changes underlying reproductive transitions in wild tomatoes
1046 (Solanum sect. Lycopersicon). Molecular Ecology. 23(8):1965-1978.

1047 WALSH NE, CHARLESWORTH D. 1992 Evolutionary interpretations of differences in pollen-tube
1048 growth rates Quarterly Review of Biology. 67(1):19-37.

1049 WILBURN DB, SWANSON WJ. 2016 From molecules to mating: Rapid evolution and biochemical
1050 studies of reproductive proteins. Journal of Proteomics. 135:12-25.

1051 WILLSON MF. 1979 Sexual selection in plants American Naturalist. 113(6):777-790.

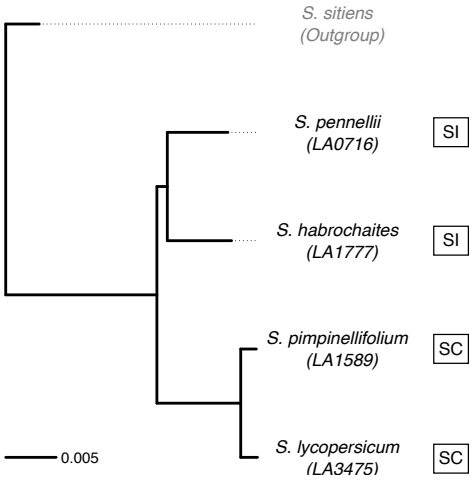
1052 WILLSON MF. 1994 Sexual selection in plants - perspective and overview American Naturalist.
1053 144:S13-S39.

1054 WONG A. 2010 Testing the effects of mating system variation on rates of molecular evolution in
1055 primates. Evolution. 64(9):2779-2785.

1056 WONG A. 2011 The molecular evolution of animal reproductive tract proteins: what have we
1057 learned from mating-system comparisons? International Journal of Evolutionary Biology.
1058 2011(908735):1-9.

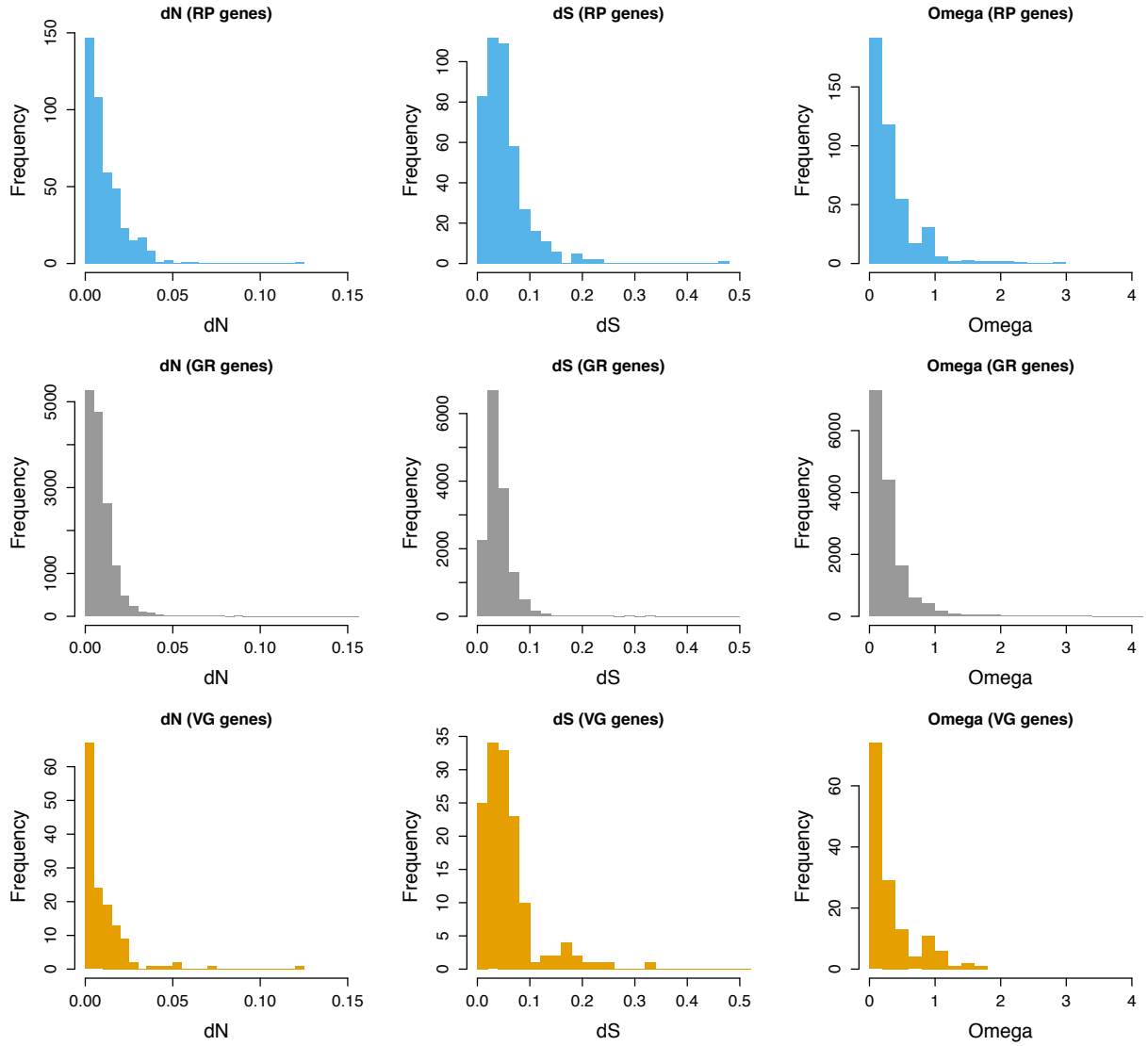
1059 YANG Z. 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and
1060 Evolution. 24(8):1586-1591.
1061
1062

1063 **Supplementary Figures**
1064
1065 **Supplementary Figure 1:** Maximum likelihood tree of focal species relationships relative to an
1066 outgroup species *Solanum sitiens*. Branch lengths (in average substitutions per site) were
1067 obtained from Pease et al. (2016a). Original tree was inferred with RAxML v.8 (Stamatakis,
1068 2014) from a whole-transcriptome concatenated alignment.
1069
1070



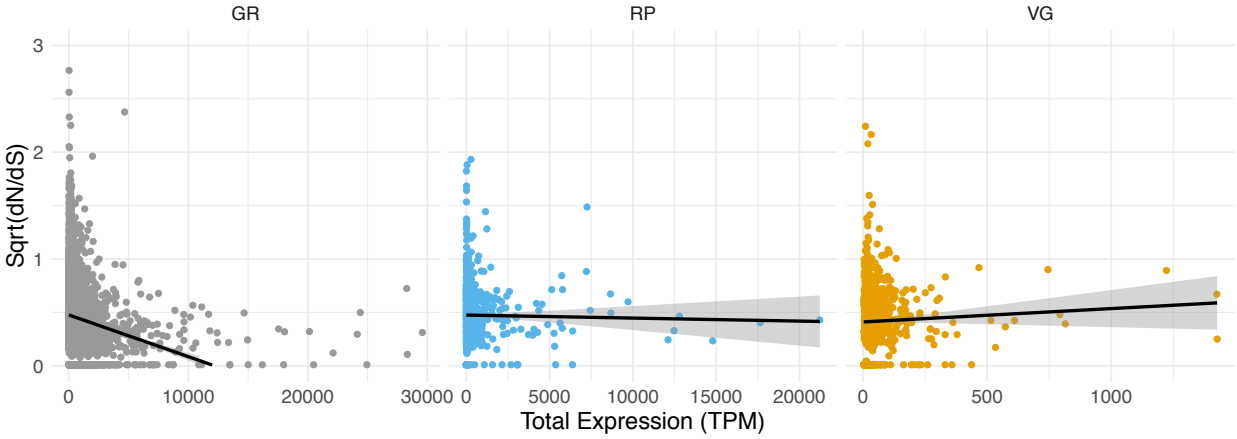
1071

1072 **Supplementary Figure 2:** Distributions of inferred per-locus dN (left), dS (middle), and dN/dS
1073 (omega) (right) from each broad class of loci: RP (upper panels), GR (middle panels), and VG
1074 (lower panels). Supporting analyses are in Table S6.
1075
1076



1077

1078 **Supplementary Figure 3:** Relationship between per locus mean level of gene expression (TPM)
1079 and per locus protein evolution (dN/dS) in loci expressed in (left) general/GR, (middle)
1080 reproductive/RP, and (right) vegetative/VG tissues in Solanum species. Each point indicates an
1081 individual locus. Regression lines shown in black; 95% CI shown in grey. Supporting analyses are
1082 in Table S13.
1083



1084
1085