

1 Article Type: Letter

2 Title: Spatial proximity determines post-speciation introgression in Solanum

3 Short Title: Geography determines post-speciation introgression

4

5

6 Jennafer A. P. Hamlin<sup>1\*</sup>

7 Leonie C. Moyle<sup>2</sup>,

8 <sup>1</sup>Department of Plant Biology, University of Georgia, 2502 Miller Plant Sciences,  
9 Athens, GA, USA, 30602

10

11 <sup>2</sup>Department of Biology, Indiana University, 1001 East Third Street, Bloomington,  
12 Indiana, USA, 47405

13

14 \*Correspondence should be sent to: [jhamlin@uga.edu](mailto:jhamlin@uga.edu)

15 Keywords: Hybridization, Geographic proximity, Mating system variation, Whole-  
16 Genome, tomato

17

18 Total Word Count: ~5304

19

20

21 ABSTRACT

22           An increasing number of phylogenomic studies have documented a clear  
23 ‘footprint’ of post-speciation introgression among closely-related species. Nonetheless,  
24 systematic genome-wide studies of factors influencing the likelihood of introgression  
25 remain rare. Here, we use an *a priori* hypothesis-testing framework, and introgression  
26 statistics, to evaluate the prevalence and frequency of introgression. Specifically, with  
27 whole genome sequences from 32 lineages of wild tomato species, we assess the  
28 effect of three factors on introgression: genetic relatedness, geographical proximity, and  
29 mating system differences. Using multiple trios within the ‘ABBA-BABA’ test, we find  
30 that one of our factors, geographic proximity, is consistently associated with evidence  
31 for recent introgression between species. Of 14 species pairs with ‘proximate’ versus  
32 ‘distant’ population comparisons, 12 showed evidence of introgression; in ten of these  
33 cases, this was more prevalent between geographically-closer populations. We found  
34 no evidence that introgression varies systematically with increasing genetic divergence  
35 between lineages or with mating system differences, although we have limited power to  
36 address the latter effect. While our analysis indicates that recent post-speciation  
37 introgression is frequent in this group, estimated levels of genetic exchange are modest  
38 (0.05-1.5% of the genome), so the relative importance of hybridization in shaping the  
39 evolutionary trajectories of these species could be limited. Regardless, similar clade-  
40 wide analyses of genomic introgression would be valuable for disentangling the major  
41 ecological, reproductive, and historical determinants of post-speciation gene flow, and  
42 for assessing the relative importance of introgression as a source of evolutionary  
43 change.

## 44 IMPACT STATEMENT

45           The formation of new species is traditionally viewed as a tree-like branching  
46 process, in which species are discrete branches that no longer share an ongoing  
47 genealogical connection with other, equally discrete, species. Recently this view has  
48 been challenged by numerous studies examining genealogical patterns across entire  
49 genomes (all the DNA of an organism); these studies suggest that the exchange of  
50 genes between different species (known as ‘introgression’) is much more common than  
51 previously appreciated. This unexpected observation raises questions about which  
52 conditions are most important in determining whether species continue to exchange  
53 genes after they diverge. Factors such as physical proximity, differences in reproductive  
54 mechanisms, and time since species shared a common ancestor, might all contribute to  
55 determining the prevalence of introgression. But to evaluate the general importance of  
56 these factors requires more than individual cases; many species comparisons, that  
57 differ systematically in one or more of these conditions, are needed. Here we use  
58 whole-genome information from 32 lineages to evaluate patterns of introgression among  
59 multiple species in a single, closely related group—the wild tomatoes of south America.  
60 We contrast these patterns among pairs of lineages that differ in their geographical  
61 proximity, reproductive system, and time since common ancestry, to assess the  
62 individual influence of each condition on the prevalence of introgression. We find that  
63 only one of our factors—geographical proximity—is consistently associated with greater  
64 evidence for recent introgression, indicating that this is largely shaped by the  
65 geographical opportunity for hybridization, rather than other plausible biological  
66 processes. Our study is one of the first to systematically assess the influence of general

- 67 ecological and evolutionary conditions on the frequency of post-speciation introgression.
- 68 It also provides a straightforward, generalizable, hypothesis-testing framework for
- 69 similar systematic analyses of introgression in groups of other organisms in the future.

## 70 INTRODUCTION

71           The prevalence of hybridization among species, and the importance of  
72 introgression for shaping species evolution, are historically contentious questions  
73 (Mallet, 2008, 2005). Although traditionally viewed to be more common among plants  
74 (Anderson, 1968; Stebbins, 1970), evidence of hybridization and introgression is  
75 emerging for an increasingly broad range of organisms (Mallet et al., 2016). Perhaps  
76 the most famous contemporary example involves Neanderthal and modern human  
77 lineages (Mallet et al., 2016), in which ~1 – 4% of Neanderthal genome is inferred to  
78 have introgressed into some human populations. Quantifying the frequency and amount  
79 of introgression is important for understanding the historical dynamics of closely-related  
80 lineages, as well as the potential sources of genetic variation that could fuel ongoing  
81 evolutionary change. For example, if sufficiently common, gene flow between species  
82 could act as a significant source of adaptive loci, as has been observed for mimicry  
83 pattern alleles in *Heliconius* butterflies (The Heliconius Genome Consortium et al.,  
84 2012). Adaptive introgression is likely to be especially important among recently  
85 diverged lineages, where the accumulation of hybrid incompatibilities is not so  
86 advanced that it prevents the exchange of unconditionally adaptive loci when lineages  
87 come into contact. Nonetheless, the clade-wide prevalence of introgression events, and  
88 therefore their relative importance in shaping the evolutionary trajectory of close  
89 relatives, is only now beginning to be assessed (Folk et al., 2018).

90           From a genomic perspective, introgression leaves a detectable ‘footprint’:  
91 introgressed regions show distinctive patterns of historical relatedness that differ from  
92 non-introgressed regions, because they are most closely related to the donor species

93 rather than the recipient genome in which they are found (Payseur and Rieseberg,  
94 2016). Accordingly, genome-wide data is ideal for characterizing the prevalence of  
95 hybridization because it provides a detailed picture of phylogenetic relationships at loci  
96 across the genome, including in genomic regions that show patterns of relatedness  
97 inconsistent with the species as a whole. Beyond the human and butterfly examples,  
98 genome-wide data has been used to infer past introgression events among species in  
99 groups as diverse as *Saccharomyces* yeast (Morales and Dujon, 2012) *Anopheles*  
100 mosquitoes (Fontaine et al., 2015), wild tomatoes (Pease et al., 2016), and *Drosophila*  
101 (Turissini and Matute, 2017). However, while revealing the extent and timing of gene  
102 flow events is interesting in individual cases, there are few tests of the generality of  
103 introgression across whole groups of closely related species, including whether it  
104 systematically varies in frequency or extent under different biological conditions.

105         Some of the factors that could influence the frequency of hybridization and  
106 subsequent introgression include phylogenetic relatedness (i.e., genetic distance),  
107 geographical proximity, and biological factors that affect the likelihood and direction of  
108 reproductive events, such as differences in mating system. In the first case,  
109 because the strength of reproductive isolation is expected to accumulate with the  
110 amount of time since lineages diverged (Coyne and Orr, 1989), more genetic exchange  
111 might be expected to occur between more closely-related species, with diminishing  
112 rates accompanying increasing lineage differentiation. Second, genetic exchange is  
113 more likely to occur among species in close geographic proximity, where they can  
114 potentially come into physical and therefore reproductive contact (Harrison, 2012).  
115 Determining the level of spatial proximity that allows gene exchange can be challenging,

116 as it likely depends upon numerous biological factors (e.g., dispersal mechanisms) and  
117 abiotic factors (e.g., physical barriers to dispersal). Nonetheless, a general expectation  
118 is that hybridization is more likely when biological or abiotic factors do not hinder contact  
119 between lineages. For example, numerous hybrid zone studies demonstrate that the  
120 proportion of individuals with mixed ancestry usually decreases with geographic  
121 distance from the hybrid zone (Harrison and Larson, 2014, 2016).

122 Third, factors that specifically influence the timing and success of reproductive  
123 events are also expected to influence the likelihood of hybridization and introgression.  
124 For example, mating system variation can influence introgression, either directly by  
125 affecting the likelihood of successful mating between species or indirectly by influencing  
126 the longer-term likelihood that introgressed loci will persist in the recipient lineage. In the  
127 first instance, mating system differences can cause predictable asymmetries in the  
128 success of initial crosses between species. This can occur either via differences in the  
129 size or shape of reproductive organs that can lead to asymmetric mechanical isolation  
130 between lineages (e.g. where outcrossing species can fertilize inbreeding species, but  
131 not vice versa; Brothers and Delph, 2017; Levin, 1978)  
132 or—especially in plants—via differences in the presence/absence of genetically  
133 determined self-incompatibility systems, whereby pollen from self-incompatible species  
134 can fertilize ovules of self-compatible species, but self-incompatible plants actively  
135 reject pollen from self-compatible species (e.g. in *Nicotiana* (Anderson and de Winton,  
136 1931), *Petunia* (Mather and Edwardes, 1943) and *Solanum* (McGuire and Rick, 1954)).  
137 In both mechanical and active-rejection cases, outcrossing species are more likely to  
138 donate alleles to more inbreeding species compared to the reciprocal direction of gene

139 flow, reducing the potential for gene flow specifically between species with unlike  
140 mating systems. Similarly, the longer-term likelihood that introgressed loci will persist in  
141 recipient lineages can vary based on the mating system of the donor and recipient  
142 lineages, because mutational load and the efficacy of selection is expected to differ  
143 between species with histories of more or less inbreeding and different effective  
144 population sizes ( $N_e$ ) (Busch, 2005; Charlesworth et al., 1990; Harris and Nielsen, 2016;  
145 Juric et al., 2016; Lande and Schemske, 1985). In particular, introgression from  
146 outbreeding to inbreeding populations should be especially disfavored both because  
147 donor alleles are expected to have stronger deleterious fitness effects (due to genetic  
148 load that can persist in outbreeders) and because the smaller  $N_e$  recipient population is  
149 less effective at disassociating these from other non-deleterious loci before they are  
150 purged (Brandvain et al., 2014; Ruhsam et al., 2011). In comparison, the exchange of  
151 alleles between lineages with similar mating systems should be less constrained by  
152 these considerations. In general, then, no matter whether affected by initial crossing  
153 differences (from mechanical or active rejection asymmetries) or differences in the  
154 historical factors determining genetic load and effective population size, gene flow  
155 between lineages that differ in their mating system might be expected to be more  
156 constrained than gene flow between lineages with similar mating systems.

157         While these factors are expected to influence the rate and likelihood of gene flow  
158 between species, there are few systematic tests of their general importance in shaping  
159 the prevalence of post-speciation introgression. Here we use whole genome data to  
160 systematically evaluate the effects of genetic distance, geographical proximity, and  
161 mating system differences, on genome-wide patterns of post-speciation introgression



162 across a closely related clade of species. To assess introgression, we use the ‘ABBA-  
163 BABA’ test (also known as the D-test; Durand et al., 2011; Green et al., 2010). This test  
164 detects introgression by comparing the frequency of alternate ancestral (“A”) and  
165 derived (“B”) allele patterns among four taxa, where the species tree has the allele  
166 pattern BBAA (Figure 1). In the absence of gene flow, the alternate minority patterns of  
167 ABBA and BABA should be approximately equally frequent, as they have an equal  
168 chance of either coalescence pattern under incomplete lineage sorting (ILS; Durand et  
169 al., 2011). In comparison, an excess of ABBA patterns indicates gene flow between  
170 lineage P2 and P3, and excess BABA indicates gene flow between lineage P1 and P3  
171 (Figure 1).

172       Importantly, the structure of the ABBA-BABA test allows us to test *a priori*  
173 hypotheses about the expected prevalence of introgression in multiple parallel  
174 comparisons. In particular, we can intentionally construct tests of a specific  
175 introgression hypothesis by consistently positioning taxa in the P1, P2, and P3 positions  
176 in a 4-taxon tree, so that P2 and P3 are always predicted to experience more  
177 introgression than P1 and P3 (Figure 1). For example, if geographic proximity *per se*  
178 affects the amount of introgression between two species, in a case where P1 and P2  
179 are populations of one species but P2 is more geographically proximate to populations  
180 of a second species (P3), then our *a priori* expectation is that elevated introgression will  
181 be observed as an excess of ABBA (i.e. evidence of gene flow between P2 and P3)  
182 rather than BABA (gene flow between P1 and P3; Figure 1). Multiple different four-taxon  
183 tests with the same structure can then be used to evaluate whether geographically  
184 proximity is systematically more frequently associated with evidence of post-speciation

185 introgression. A similar structure can be used to test the *a priori* expectation that gene  
186 flow is expected to be more frequent between lineages with shared versus different  
187 mating systems. More generally, 4-taxon tests that involve increasing evolutionary  
188 divergence between the P1/P2 and P3 lineages, can be used to evaluate evidence that  
189 introgression is on average more prevalent between more closely related taxa.

190       Here we use this *a priori* hypothesis-testing framework to assess the prevalence  
191 and frequency of introgression among wild tomato lineages (*Solanum* section  
192 *Lycopersicum*) depending upon a) geographical proximity, b) differences in mating  
193 system, and c) evolutionary distance. In addition to high-quality, whole-genome data for  
194 multiple wild genotypes of known provenance (Aflitos et al., 2014; Lin et al., 2014), this  
195 group has well-curated historical records of geographic coordinates for hundreds of wild  
196 populations from across the clade. Moreover, prior evidence of introgression events  
197 between specific lineages (e.g. Beddows et al., 2017; Pease et al., 2016), and the ability  
198 to generate F1 and later-generation hybrids in the greenhouse (e.g., Moyle, 2008; Rick,  
199 1979), indicate the possibility that introgression could shape genomes in this group.  
200 Using whole-genome data from 32 closely related accessions across 12 species of wild  
201 tomato, our goal here was to systematically test hypotheses about the prevalence of  
202 introgression, to make general inferences about the role and importance of particular  
203 factors in the frequency of cross-species hybridization, and to begin to assess the  
204 potential importance of introgression in shaping genome content and evolutionary  
205 trajectories in this clade.

## 206 METHODS

### 207 *Sequencing data and mapping to reference genome*

208 Our analyses used data from three whole genome-sequencing projects (Aflitos et  
209 al., 2014; Hardigan et al., 2016; Lin et al., 2014). Specifically, we obtained raw reads  
210 either as fastq or SRA files for genomes of 32 wild *Solanum* individuals from the tomato  
211 clade (*Solanum* section *Lycopersicum*), each from a different accession (historical  
212 population collection), along with *Solanum tuberosum* (potato; Hardigan et al., 2016)  
213 which we used as the outgroup in all comparisons (Supplemental Table 1). To combine  
214 data from the different sequencing projects, we trimmed and re-mapped raw reads back  
215 to the reference genome of domesticated tomato, *S. lycopersicum* version 2.50 (The  
216 Tomato Genome Consortium, 2012), using standard practices for mapping and quality-  
217 filtering (see Supplemental Text).

218

### 219 *Hypothesis testing with the D-statistic*

220 We used the ABBA-BABA test to assess evidence for the presence and  
221 directionality of gene flow in a set of 4-taxon tests. The results of each ABBA-BABA test  
222 can be expressed in terms of Patterson's *D*-statistic, calculated as  $(\#ABBA - \#BABA) /$   
223  $(\#ABBA + \#BABA)$  for all biallelic sites in the multiple sequence alignment (Durand et  
224 al., 2011; Green et al., 2010). The *D*-statistic therefore summarizes both the magnitude  
225 of introgression and the specific pair of taxa that are exchanging alleles; positive values  
226 of *D* indicate P2 and P3 are exchanging more alleles (an excess of ABBA) and negative  
227 values indicate more gene exchange between P1 and P3 (an excess of BABA). We  
228 used multiple replicate 4-taxon tests to evaluate three *a priori* expectations:

229 1) Post-speciation introgression is more prevalent between geographically closer  
230 versus more distant lineages. Four-taxon tests were structured so that P1 and P2 were

231 taken from populations of the same species, but P2 was spatially closer to the P3  
232 species and P1 was more distant (Figure 1). In this case, we expect a systematic  
233 excess of positive values of the resulting *D*-statistics. Species comparisons and specific  
234 P1, P2, and P3 accessions were identified for these tests based on known species  
235 ranges and geographical locations of the sequenced accessions (see Supplemental text  
236 for our specific criteria). Because these analyses are constrained by the available  
237 sequenced genotypes, the actual geographic distances involved vary broadly between  
238 4-taxon tests (Supplemental Table 2), so that this analysis is an imperfect reflection of  
239 close spatial proximity; however, the structure of each test means we are still able to  
240 systematically compare the effect of greater ('proximate') versus less ('distant')  
241 geographic proximity on detected patterns of introgression.

242         2) Post-speciation introgression is more prevalent between lineages that share  
243 mating system. Here, four-taxon tests were structured so that P1 and P2 were again  
244 taken from populations of the same species, but P2 was self-compatible (SC) and P1  
245 was self-incompatible (SI); in every test, P3 was the same accession of an SC species.  
246 In this case, we also expect a systematic excess of positive *D*-statistics. Within our  
247 dataset there are only three species for which we had whole genome sequence data  
248 from both SI and SC accessions: *S. arcanum*, *S. habrochaites*, and *S. peruvianum*.

249         3) Post-speciation introgression is more prevalent between lineages that are  
250 more closely related. In this case, we expect that the estimated magnitude of *D* should  
251 decrease as evolutionary (genetic) distance between (P1, P2) and P3 increases within  
252 the 4-taxon test. We calculated pairwise genetic distance for each comparison by taking  
253 the average genetic distance for the two comparisons within the focal trio (i.e. P1 with

254 P3 and P2 with P3), based on genome-wide site differences between accessions. In  
255 this case, there is no expectation of either positive or negative  $D$ -values, as P1 and P2  
256 are equally closely related to P3. Instead, the comparison is among  $D$ -values taken from  
257 different 4-taxon combinations.

258 The supplemental text provides a more detailed description of how these factors  
259 were defined and determined for individual 4-taxon tests.

260

### 261 *Calculating and analyzing D-statistics*

262 Using the program *mvftools* (Pease and Rosenzweig, 2018), we estimated  $D$ -  
263 statistics for consecutive 100kb windows across each of the 12 unique chromosomes of  
264 *Solanum*, and then calculated the genome-wide average of  $D$  across all of these  
265 windows. 100 kb windows were defined based on the domesticated tomato genome,  
266 resulting in 8,015,000 total windows (0.8015 gigabases or ~ 80% of the genome). For all  
267 analyses reported in the main text, we used only 100 kb windows that had >20 variable  
268 sites (~4200 windows on average per test, depending upon the specific trio analyzed;  
269 Supplemental Table 3); this eliminated windows in which no or few single nucleotide  
270 polymorphisms (SNPs) were found, as these have low power to accurately estimate  $D$ .  
271 For completeness, we also performed analyses on  $D$  estimated from all 100 kb windows  
272 in the genome and found the same qualitative results as the higher power dataset  
273 (Supplemental Table 4). For each four-taxon test, we performed a chi-squared  
274 goodness-of-fit test, to determine if our genome-wide estimate of mean  $D$  was  
275 significantly different than zero. Across all four-taxon tests, we determined whether the  
276 number of tests supporting a higher incidence of introgression in the predicted direction

277 was greater than expected by chance, using a sign test. All statistical analyses were  
278 performed in RStudio version 1.1453 (2015).

279 To determine if post-speciation introgression is associated with quantitative  
280 differences in geographical proximity between populations of two species, for every trio  
281 we also determined the geographic distance (in km) between P1 and its nearest P3  
282 accession, and between P2 and its nearest P3 accession (using our georeferenced  
283 location data for all population accessions, see supplemental text), to generate an  
284 estimate of their relative proximity to any population from P3 (i.e. the difference between  
285 these two distances; Supplemental Table 5). For all four-taxon tests, we regressed our  
286 mean genome-wide estimate of  $D$  on this relative geographic distance estimate  
287 (Supplemental Figure 1). To determine if evidence for post-speciation introgression is  
288 more prevalent for recently diverged lineages, we evaluated the association between  
289 the absolute values of our estimates of genome-wide  $D$  and genome-wide genetic  
290 distance, using regression across all four-taxon tests.

291 Finally, to provide a general estimate of the fraction of the genome inferred to  
292 have come from introgression in each trio, we calculated the number of varying sites  
293 that support evidence of differential introgression (i.e. the excess of sites that support  
294 one minority topology over the other, or  $(|ABBA \text{ sites} - BABA \text{ sites}|)$ ) and expressed  
295 these as a proportion of total variable sites in this trio (i.e.  $|ABBA - BABA| / (ABBA +$   
296  $BABA + BBAA)$ ). These estimates were calculated for all four-taxon combinations  
297 included in our analyses, and used only data from 100 kb windows with  $>20$  SNPs  
298 (Supplemental Table 6). While these estimates are not precise measures of the  
299 proportion of the genome that has experienced introgression (for example, they could

300 underestimate total introgression if there has been gene flow both between P1 and P3,  
301 and between P2 and P3), they provide a set of rough global estimates with which to  
302 compare different four-taxon tests. These estimates were regressed onto mean genetic  
303 distance for all trios to evaluate whether the amount of inferred introgression between  
304 species was associated with the time since their divergence (Supplemental Figure 2).

305

## 306 RESULTS

307 We generated genome-wide  $D$  estimates for 17 four-taxon combinations, 14 that  
308 could address the effects of geographical proximity on introgression, and three  
309 addressing the effects of mating system variation. All 17 combinations were used to  
310 assess the effect of genetic distance. Across all 17 tests, the genome-wide average  
311 estimates of  $D$  ranged from -0.08 to 0.15, based on 100kb windows with >20 variable  
312 sites (Table 1; estimates of  $D$  from all windows were similar, Supplemental Table 4).  
313 The fraction of the genome estimated to be differentially introgressed ranged from 0.05  
314 – 1.5 (Supplemental Table 6).

315 For our tests of geographical proximity, of 14 testable four-taxon combinations,  
316 ten had average  $D$  values significantly greater than zero—indicating that our  
317 geographically closer lineages (P2 and P3) share a higher proportion of sites—whereas  
318 two were significantly less than zero (chi-squared goodness-of-fit tests, with Bonferroni  
319 correction; Figure 2, Table 1a). A two-sided sign-test indicated that the number of  
320 significantly positive versus negative mean values of  $D$  was different ( $p = 0.038$ ),  
321 consistent with a systematic excess of introgression between species pairs when their  
322 populations are geographically closer versus more distant. Because there is non-

323 independence in our dataset (that is, some individual accessions/genome sequences  
324 are used in more than one four-taxon test), we evaluated the influence of this non-  
325 independence by paring our dataset down to trios (of P1, P2, P3) that only used unique  
326 accessions. Of our twelve trios with significant positive or negative  $D$  statistics, we could  
327 evaluate combinations of up to six different trios that shared no accessions in common.  
328 In all eight unique alternative combinations of six trios, five tests had a positive  $D$  and  
329 one had a negative  $D$ ; that is, we found evidence for an excess of introgression  
330 consistently more frequently when populations of a species pair were geographically  
331 closer versus more distant. This directionality is non-significant in each of the reduced  
332 datasets ( $p = 0.2188$ ; Supplemental Table 7) as the two sided sign test is  
333 underpowered to detect a systematic difference in direction when  $n=6$ . Finally, across all  
334 14 four-taxon tests, mean genome-wide  $D$  was not significantly associated with the  
335 relative geographical proximity of the P1 versus P2 population to the closest population  
336 from the P3 species ( $R$ -squared = 0.055,  $p$ -value = 0.42, Supplemental Figure 1a),  
337 suggesting that, while geographical proximity influences the possibility of introgression,  
338 other factors likely determine the quantitative amount of introgression that occurs.

339 For our evaluation of mating system effects, of three testable trios, two genome-  
340 wide mean  $D$  values were significantly positive and one was significantly negative,  
341 regardless of the specific accession used as the P3 taxon in our tests (Table 1; Figure  
342 3). Note that relative geographical proximity does not explain the sign of  $D$  in any of  
343 these cases; for the two instances where  $D$  was positive (P1 and P2 from *S. arcanum* or  
344 from *S. habrochaites*, respectively: Supplemental Table 5), our SI accessions are  
345 geographically closer to a P3 accession compared to their conspecific SC accessions,



346 and the opposite is observed in the case where mean  $D$  was negative. In addition,  
347 mean genome-wide  $D$  is not negatively associated with the relative difference in  
348 geographic proximity for these mating system trios (Supplemental Figure 1b). Overall,  
349 then, our evidence in support of mating system effects is equivocal: 2 of 3 tests support  
350 the expectation that introgression is more prevalent between species pairs when their  
351 populations share the same mating system, but one does not.

352 Finally, we detected no association between quantitative values of  $D$  and  
353 increasing evolutionary divergence between P1/P2 and P3 species (for 14 geographic  
354 trios:  $R$ -squared = 0.021,  $p$ -value = 0.62; for all 17 trios:  $R$ -squared = 0.012,  $p$ -value =  
355 0.68, Figure 4 and Supplemental Table 8), suggesting little evidence that the propensity  
356 for introgression is determined by variation in evolutionary divergence among the  
357 species pairs examined here. Similarly, the estimated amount of differential  
358 introgression (as a proportion of all variable sites; 0.05%-1.5%) is unrelated to the mean  
359 genetic distance between the focal species in each trio ( $R$ -squared = 0.0004,  $p$ -value =  
360 0.938, Supplemental Figure 2).

361

## 362 DISCUSSION

363 The prevalence of introgression is one pattern emerging from contemporary  
364 genome-wide studies in many groups of closely related species, including in groups not  
365 traditionally associated with post-speciation gene flow. However, there have been few  
366 attempts to systematically assess the influence of different factors in shaping the  
367 frequency and extent of this gene flow. Here we used directionally structured, four-taxon  
368 ABBA-BABA tests to examine the influence of three factors—genetic distance,

369 geographical proximity, and mating system differences—on genome-wide patterns of  
370 introgression among wild tomato species. We found that recent introgression was  
371 commonly detected among these species, that consistent patterns of post-speciation  
372 introgression depend largely on geographical proximity rather than the other two factors,  
373 and that the estimated fraction of the genome differentially introgressed between  
374 species was modest. These findings have interesting implications for interpreting the  
375 contexts in which introgression might play the greatest role in shaping evolutionary  
376 trajectories in this and other similar clades, and for assessing the potential contribution  
377 of introgression to adaptive phenotypic evolution.

378

379 *Recent introgression occurs frequently but is modest in scope among wild tomatoes*

380 Our analysis indicates that, among wild tomato lineages, post-speciation gene  
381 exchange is prevalent: of 17 total four-taxon tests across all our analyses, 15 had mean  
382  $D$  values significantly different from zero. Prior studies have detected evidence for  
383 introgression among specific wild tomato lineages (Beddows et al., 2017; Pease et al.,  
384 2016), and our findings expand and illuminate these observations in several key  
385 respects. First, our analyses preferentially assessed evidence for recent, rather than  
386 more ancient (Pease et al., 2016), introgression events because in every case we  
387 contrasted populations (P1 and P2) from a single species when looking for evidence of  
388 introgression with a second species (P3). Accordingly, any inferred introgression must  
389 have occurred after the evolutionary split of these two (P1 and P2) conspecific  
390 populations. Despite this, we find repeated evidence that populations from different  
391 species have exchanged genes recently, including species that are estimated to have

392 diverged >2 million years ago (e.g., *S. pimpinellifolium* and *S. pennellii*; Pease et al.,  
393 2016).

394 Our results suggest there is broad potential for cross-species hybridization  
395 across the clade, a finding consistent with other observations that indicate premating  
396 isolation is likely to be incomplete between lineages in nature: all species share general  
397 floral morphology (rotate, yellow, five-petaled flowers), all are buzz-pollinated, and  
398 multiple species pairs are found in sympatry (Rick, 1950). Nonetheless this finding is  
399 intriguing as few natural hybrids have been observed in the wild in this group (Taylor,  
400 1986), and some of these species are known via crossing and genetic studies to  
401 express moderate to strong postmating and postzygotic reproductive isolation under lab  
402 conditions (Hamlin et al., 2017; Moyle and Nakazato, 2010, 2008). These later-acting  
403 barriers might be important in limiting the amount of introgression that results from  
404 hybridization events. Indeed, a second general observation of our analysis is that  
405 despite evidence for relatively frequent hybridization, the amount of the genome  
406 exchanged between species is likely to be limited: the proportion of the genome  
407 estimated to be differentially exchanged between species is on the order of 0.05-1.5%  
408 of all variable sites (Supplementary Table 8). Further, when D-statistics are examined  
409 chromosome-by-chromosome within each four-taxon test (Supplementary Figure 3), in  
410 most cases introgression is inferred on some chromosomes but not others; this variation  
411 among chromosomes might be due to variation in the presence of loci contributing to  
412 reproductive isolation. Overall, the amount and distribution of inferred introgression  
413 suggests that current species reproductive barriers are sufficiently incomplete to allow  
414 detectable recent introgression among diverged species in the field, but also that

415 genomes are not completely or uniformly porous to gene flow among lineages, even in  
416 cases where there is an opportunity for gene exchange.

417

418 *Introgression frequency varies with spatial proximity between species pairs,*  
419 *rather than overall genetic relatedness or mating system differences*

420 Importantly, the structure of our analyses also allowed us to explicitly evaluate  
421 the influence of several factors on these detected patterns of introgression. We found  
422 that repeated patterns of recent post-speciation hybridization were consistently  
423 associated with only one of our factors: geographical proximity. This suggests that the  
424 propensity for post-speciation gene exchange is most often dependent on the simple  
425 opportunity for reproductive contact—geographic proximity *per se* between populations  
426 of different species allows greater gene flow. In contrast, although there were few  
427 comparisons available, we found that mating system differences were not consistently  
428 associated with reduced (or elevated) patterns of introgression; two tests were  
429 consistent with our *a priori* prediction (including the strongest case) but one test  
430 indicated more introgression between lineages that differed in mating system. While we  
431 currently have too few four-taxon tests to draw definitive conclusions, these  
432 idiosyncratic observations could be due to additional biological factors. For example, the  
433 transition to self-compatibility within each of our polymorphic (SI/SC) species might be  
434 too recent to yet see consistent differences in crossing asymmetry or genetic load that  
435 are expected between long-standing SC versus SI populations. Regardless, because  
436 there are clear predictions about associations between mating system, genetic load, the  
437 efficacy of selection, and the propensity and direction of introgression (Busch, 2005;

438 Charlesworth et al., 1990; Harris and Nielsen, 2016; Juric et al., 2016; Lande and  
439 Schemske, 1985)—some of which appear to be supported in individual cases (e.g.,  
440 *Mimulus*; Brandvain et al., 2014)—testing the generality of these effects with a larger set  
441 of comparisons remains a goal in the future. Within the limitations of the data available  
442 here, however, mating system difference alone is an equivocal predictor of overall  
443 patterns of recent genomic introgression in this group.

444         Interestingly, we also did not detect an association between the magnitude of  
445 evolutionary divergence (genetic distance), and either the occurrence or the amount of  
446 inferred introgression. Species are expected to accumulate reproductive isolation with  
447 increasing evolutionary divergence (Coyne and Orr, 1997), and this pattern has been  
448 observed among wild tomatoes for loci involved in postzygotic reproductive isolation  
449 (hybrid pollen and seed sterility; Moyle and Nakazato, 2010), suggesting that  
450 introgression should become attenuated with increasing evolutionary age between  
451 species. Nonetheless, the total number of loci estimated to contribute to postzygotic  
452 isolation in this group is relatively modest, even among the oldest species pairs (Moyle  
453 and Nakazato, 2010, 2008). In addition, mean sequence divergence between all  
454 lineages analyzed here is low—0.1-4.0%—consistent with the recent, rapid origin of  
455 species in this clade (Pease et al., 2016). Using data from a very broad range of taxa, a  
456 recent meta-analysis inferred that divergence of just a few percent results in barriers  
457 that can effectively suppress gene flow (Roux et al., 2016) indicating that genetic  
458 divergence can be a strong determinant of introgression, at least beyond some  
459 threshold at which isolating barriers are sufficiently strong. The data presented here  
460 suggest that wild tomato species have not yet exceeded this threshold. Instead, our

461 inference is that it is largely the opportunity for gene exchange, via geographical  
462 proximity, that currently determines the possibility of post-speciation introgression in this  
463 group.

464         Finally, as this inference suggests, the conditions that generally favor or prevent  
465 gene flow between species will likely vary depending on the biological features of  
466 different systems. Although there are yet too few comprehensive tests to determine how  
467 much they might vary, some previous analyses have assessed factors influencing  
468 general introgression patterns with different approaches. For example, Winger (2017)  
469 evaluated the relationship between introgression and plumage differentiation for 16  
470 lineages of Andean cloud forest birds within a geographically and ecologically structured  
471 study. He found evidence for introgression across a geographic barrier between lineage  
472 pairs with uniform plumage patterns, but not between pairs with divergent plumage. This  
473 suggests that different patterns of sexual selection might determine whether and when  
474 introgression is expected, although alternative explanations, including more time since  
475 divergence between plumage-differentiated pairs, could not be excluded in this case.  
476 Clearly, additional systematic tests of the ecological, reproductive, and historical factors  
477 most strongly predictive of post-speciation gene flow will be helpful in evaluating how  
478 these might or might not differ between major groups of organisms.

479

#### 480 *Evaluating introgression as an important evolutionary force*

481         Our analyses join a growing consensus of studies that suggest gene flow among  
482 distinct lineages might be common, especially those that have rapidly diverged, have  
483 incomplete isolating barriers, and that maintain some regions of geographical overlap

484 (Brawand et al., 2014; Jónsson et al., 2014; Lamichhaney et al., 2015; The Heliconius  
485 Genome Consortium et al., 2012). What do these observations indicate about whether  
486 introgression contributes substantially to shaping the evolution of these lineages? The  
487 expectations here are complex. When gene flow is frequent the opportunity for  
488 introgressed alleles to contribute to evolutionary responses in the recipient lineage  
489 clearly increases, but gene flow could also have deleterious consequences by  
490 introducing alleles that are maladaptive in this new background. Instead, large amounts  
491 of gene flow between species might indicate that introgressing loci have little or no  
492 detectable fitness effects: that is—that they are adaptively inconsequential. Conversely,  
493 even very restricted gene flow might still be consistent with the possibility of ‘adaptive’  
494 introgression—the movement of alleles between species that increase fitness in their  
495 new recipient lineage (Suarez-Gonzalez et al., 2018). Several of the best cases of  
496 apparently adaptive introgression involve the movement of relatively small chromosomal  
497 segments (e.g. mimicry loci in *Heliconius*; high altitude adaptation in ancestral human  
498 populations; Suarez-Gonzalez et al., 2018; The Heliconius Genome Consortium et al.,  
499 2012). Even a genetic background of potentially deleterious interactions (that cause  
500 reduced hybrid fitness) a fairly limited exchange of alleles might be expected, especially  
501 if the associated variants have very large but local fitness consequences, as is  
502 observed for locally adaptive mimicry alleles in *Heliconius* for example. While not a  
503 primary goal of our analyses here, we did not observe any obvious analogous cases at  
504 mating-system loci in our genomes (see Supplementary Text). A prior analysis using  
505 transcriptome data did detect limited regions of introgression among some *Solanum*  
506 species pairs, with indirect evidence that these might have adaptive functions

507 associated with local ecological conditions [0.0] definitely demonstrating adaptive  
508 function requires knowing the fitness effects of alternative alleles under ecologically  
509 realistic conditions, these remain hypotheses. For now, the observation that a low level  
510 of gene exchange frequently occurs between lineages suggests that introgression could  
511 be a significant source of adaptive genetic variation, certainly in comparison to lineages  
512 where there is no evidence of gene flow.

513 Overall, while genetic exchange in any particular instance is likely to be  
514 influenced by both ecological and genomic contexts, only via systematic tests of  
515 introgression patterns across multiple cases can we start to disentangle the major  
516 determinants of post-speciation gene flow. One of our goals here was to demonstrate  
517 that structured *a priori* tests, across multiple species pairs in a diverse clade, provides  
518 one method for assessing the relative influence of several general factors on the  
519 frequency and amount of post-speciation introgression. By going beyond individual  
520 cases to generate general and quantitative evaluations, these tests should provide  
521 insight both into the main determinants of post-speciation introgression across a  
522 diversity of organisms and contexts, and the relative importance of introgression as an  
523 engine of evolution.

524



525 ACKNOWLEDGEMENTS

526           The authors thank members of the Moyle Lab and the Hahn Lab (Indiana  
527 University) for advice on genomic analyses, in addition to M. Behringer, J. Pease, and  
528 M. Hahn. The authors would also like to thank # anonymous reviewers for comments  
529 that greatly improved the manuscript. This research was funded by National Science  
530 Foundation grant MCB-1127059.

531

532 AUTHOR CONTRIBUTIONS

533           J.A.P.H. and L.C.M. designed the experiments; J.A.P.H. conducted the  
534 bioinformatic analyses; J.A.P.H. and L.C.M. wrote the paper.

535

536 DATA ACCESSIBILITY

537           Data used in the analyses are available from the NCBI SRA (Wild accessions of  
538 *Solanum* section *Lycopersicum*: SRP045767; *Solanum tuberosum*: SRP059592) or the  
539 European Nucleotide Archive (PRJEB5235). Output files generated by mvftools  
540 available via data dryad (XXXX).

541 LITERATURE CITED

542

543 Aflitos, S., Schijlen, E., Jong, H. de, Ridder, D. de, Smit, S., Finkers, R., Wang, J., Zhang, G.,  
544 Li, N., Mao, L., Bakker, F., Dirks, R., Breit, T., Gravendeel, B., Huits, H., Struss, D.,  
545 Swanson □ Wagner, R., Leeuwen, H. van, Ham, R.C.H.J. van, Fito, L., Guignier, L.,  
546 Sevilla, M., Ellul, P., Ganko, E., Kapur, A., Reclus, E., Geus, B. de, Geest, H. van de,  
547 Hekkert, B. te L., Haarst, J. van, Smits, L., Koops, A., Sanchez □ Perez, G., Heusden,  
548 A.W. van, Visser, R., Quan, Z., Min, J., Liao, L., Wang, X., Wang, G., Yue, Z., Yang, X.,  
549 Xu, N., Schranz, E., Smets, E., Vos, R., Rauwerda, J., Ursem, R., Schuit, C., Kerns, M.,  
550 Berg, J. van den, Vriezen, W., Janssen, A., Datema, E., Jahrman, T., Moquet, F.,  
551 Bonnet, J., Peters, S., 2014. Exploring genetic variation in the tomato (*Solanum section*  
552 *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148.  
553 <https://doi.org/10.1111/tpj.12616>

554 Anderson, E., 1949. *Introgressive hybridization*. John Wiley and Sons, New York, NY.

555 Anderson, E., de Winton, D., 1931. The Genetic Analysis of An Unusual Relationship between  
556 Self-Sterility and Self-Fertility in *Nicotiana*. *Ann. Mo. Bot. Gard.* 18, 97–116.  
557 <https://doi.org/10.2307/2394046>

558 Beddows, I., Reddy, A., Kloesges, T., Rose, L.E., 2017. Population Genomics in Wild  
559 Tomatoes—The Interplay of Divergence and Admixture. *Genome Biol. Evol.* 9, 3023–  
560 3038. <https://doi.org/10.1093/gbe/evx224>

561 Brandvain, Y., Kenney, A.M., Flagel, L., Coop, G., Sweigart, A.L., 2014. Speciation and  
562 Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet.* 10,  
563 e1004410. <https://doi.org/10.1371/journal.pgen.1004410>

564 Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y.,  
565 Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P.,  
566 Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.-F., Barloy-Hubler, F., Berlin,  
567 A., Bloomquist, R., Carleton, K.L., Conte, M.A., D’Cotta, H., Eshel, O., Gaffney, L.,  
568 Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W.,  
569 Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P.,  
570 MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J.,  
571 Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W.,  
572 Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J.,  
573 Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S.,  
574 Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh,  
575 K., Seehausen, O., Di Palma, F., 2014. The genomic substrate for adaptive radiation in  
576 African cichlid fish. *Nature* 513, 375–381. <https://doi.org/10.1038/nature13726>

577 Brothers, A.N., Delph, L.F., 2017. Divergence in style length and pollen size leads to a  
578 postmating-prezygotic reproductive barrier among populations of *Silene latifolia*.  
579 *Evolution* 71, 1532–1540. <https://doi.org/10.1111/evo.13248>

580 Busch, J.W., 2005. The evolution of self-compatibility in geographically peripheral populations  
581 of *Leavenworthia alabamica* (Brassicaceae). *Am. J. Bot.* 92, 1503–1512.  
582 <https://doi.org/10.3732/ajb.92.9.1503>

583 Charlesworth, D., Morgan, M.T., Charlesworth, B., 1990. Inbreeding Depression, Genetic  
584 Load, and the Evolution of Outcrossing Rates in a Multilocus System With No Linkage.  
585 *Evolution* 44, 1469–1489. <https://doi.org/10.1111/j.1558-5646.1990.tb03839.x>

- 586 Coyne, J.A., Orr, H.A., 1997. "Patterns of Speciation in *Drosophila*" Revisited. *Evolution* 51,  
587 295–303. <https://doi.org/10.1111/j.1558-5646.1997.tb02412.x>
- 588 Coyne, J.A., Orr, H.A., 1989. Patterns of Speciation in *Drosophila*. *Evolution* 43, 362–381.  
589 <https://doi.org/10.1111/j.1558-5646.1989.tb04233.x>
- 590 Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for Ancient Admixture  
591 between Closely Related Populations. *Mol. Biol. Evol.* 28, 2239–2252.  
592 <https://doi.org/10.1093/molbev/msr048>
- 593 Folk, R.A., Soltis, P.S., Soltis, D.E., Guralnick, R., 2018. New prospects in the detection and  
594 comparative analysis of hybridization in the tree of life. *Am. J. Bot.* 105, 364–375.  
595 <https://doi.org/10.1002/ajb2.1018>
- 596 Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V.,  
597 Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., Mitchell, S.N., Wu, Y.-C., Smith, H.A.,  
598 Love, R.R., Lawniczak, M.K., Slotman, M.A., Emrich, S.J., Hahn, M.W., Besansky, N.J.,  
599 2015. Mosquito genomics. Extensive introgression in a malaria vector species complex  
600 revealed by phylogenomics. *Science* 347, 1258524.  
601 <https://doi.org/10.1126/science.1258524>
- 602 Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li,  
603 H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen,  
604 J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M.,  
605 Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M.,  
606 Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm,  
607 M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušić, I., Doronichev, V.B.,  
608 Golovanova, L.V., Lalueza-Fox, C., Rasilla, M. de la, Fortea, J., Rosas, A., Schmitz,  
609 R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M.,  
610 Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A Draft Sequence of  
611 the Neandertal Genome. *Science* 328, 710–722.  
612 <https://doi.org/10.1126/science.1188021>
- 613 Hamlin, J.A.P., Sherman, N.A., Moyle, L.C., 2017. Two Loci Contribute Epistatically to  
614 Heterospecific Pollen Rejection, a Postmating Isolating Barrier Between Species. *G3*  
615 *GenesGenomesGenetics* 7, 2151–2159. <https://doi.org/10.1534/g3.117.041673>
- 616 Hardigan, M.A., Crisovan, E., Hamilton, J.P., Kim, J., Laimbeer, P., Leisner, C.P., Manrique-  
617 Carpintero, N.C., Newton, L., Pham, G.M., Vaillancourt, B., Yang, X., Zeng, Z.,  
618 Douches, D.S., Jiang, J., Veilleux, R.E., Buell, C.R., 2016. Genome Reduction  
619 Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation  
620 in Asexually Propagated *Solanum tuberosum*[OPEN]. *Plant Cell* 28, 388–405.  
621 <https://doi.org/10.1105/tpc.15.00538>
- 622 Harris, K., Nielsen, R., 2016. The Genetic Cost of Neanderthal Introgression. *Genetics* 203,  
623 881–891. <https://doi.org/10.1534/genetics.116.186890>
- 624 Harrison, R.G., 2012. The Language of Speciation. *Evolution* 66, 3643–3657.  
625 <https://doi.org/10.1111/j.1558-5646.2012.01785.x>
- 626 Harrison, R.G., Larson, E.L., 2016. Heterogeneous genome divergence, differential  
627 introgression, and the origin and structure of hybrid zones. *Mol. Ecol.* 25, 2454–2466.  
628 <https://doi.org/10.1111/mec.13582>
- 629 Harrison, R.G., Larson, E.L., 2014. Hybridization, Introgression, and the Nature of Species  
630 Boundaries. *J. Hered.* 105, 795–809. <https://doi.org/10.1093/jhered/esu033>

- 631 Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M.,  
632 Albrechtsen, A., Petersen, B., Korneliussen, T.S., Vilstrup, J.T., Lear, T., Myka, J.L.,  
633 Lundquist, J., Miller, D.C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S.,  
634 Stagegaard, J., Strauss, G., Bertelsen, M.F., Sicheritz-Ponten, T., Antczak, D.F., Bailey,  
635 E., Nielsen, R., Willerslev, E., Orlando, L., 2014. Speciation with gene flow in equids  
636 despite extensive chromosomal plasticity. *Proc. Natl. Acad. Sci.* 111, 18655–18660.  
637 <https://doi.org/10.1073/pnas.1412627111>
- 638 Juric, I., Aeschbacher, S., Coop, G., 2016. The Strength of Selection against Neanderthal  
639 Introgression. *PLOS Genet.* 12, e1006340.  
640 <https://doi.org/10.1371/journal.pgen.1006340>
- 641 Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A.,  
642 Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B.R., Grant, P.R., Webster,  
643 M.T., Andersson, L., 2015. Evolution of Darwin’s finches and their beaks revealed by  
644 genome sequencing. *Nature* 518, 371–375. <https://doi.org/10.1038/nature14181>
- 645 Lande, R., Schemske, D.W., 1985. The Evolution of Self-Fertilization and Inbreeding  
646 Depression in Plants. I. Genetic Models. *Evolution* 39, 24–40.  
647 <https://doi.org/10.2307/2408514>
- 648 Levin, D.A., 1978. The Origin of Isolating Mechanisms in Flowering Plants, in: Hecht, M.K.,  
649 Steere, W.C., Wallace, B. (Eds.), *Evolutionary Biology*, Evolutionary Biology. Springer  
650 US, Boston, MA, pp. 185–317. [https://doi.org/10.1007/978-1-4615-6956-5\\_4](https://doi.org/10.1007/978-1-4615-6956-5_4)
- 651 Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X.,  
652 Huang, Z., Li, Junming, Zhang, C., Wang, T., Zhang, Yuyang, Wang, A., Zhang,  
653 Yancong, Lin, K., Li, C., Xiong, G., Xue, Y., Mazzucato, A., Causse, M., Fei, Z.,  
654 Giovannoni, J.J., Chetelat, R.T., Zamir, D., Städler, T., Li, Jingfu, Ye, Z., Du, Y., Huang,  
655 S., 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat.*  
656 *Genet.* 46, 1220–1226. <https://doi.org/10.1038/ng.3117>
- 657 Mallet, J., 2008. Hybridization, ecological races and the nature of species: empirical evidence  
658 for the ease of speciation. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 2971–2986.  
659 <https://doi.org/10.1098/rstb.2008.0081>
- 660 Mallet, J., 2005. Hybridization as an invasion of the genome. *Trends Ecol. Evol.*, Special issue:  
661 Invasions, guest edited by Michael E. Hochberg and Nicholas J. Gotelli 20, 229–237.  
662 <https://doi.org/10.1016/j.tree.2005.02.010>
- 663 Mallet, J., Besansky, N., Hahn, M.W., 2016. How reticulated are species? *BioEssays News*  
664 *Rev. Mol. Cell. Dev. Biol.* 38, 140–149. <https://doi.org/10.1002/bies.201500149>
- 665 Mather, K., Edwardes, P.M.J., 1943. Specific differences in *Petunia*. *J. Genet.* 45, 243–260.  
666 <https://doi.org/10.1007/BF02982967>
- 667 McGuire, D., Rick, C., 1954. Self-incompatibility in species of *Lycopersicon* Sect. *Eriopersicon*  
668 and hybrids with *L. esculentum*. *Hilgardia* 23, 101–124.
- 669 Morales, L., Dujon, B., 2012. Evolutionary role of interspecies hybridization and genetic  
670 exchanges in yeasts. *Microbiol. Mol. Biol. Rev.* MMBR 76, 721–739.  
671 <https://doi.org/10.1128/MMBR.00022-12>
- 672 Moyle, L.C., 2008. Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect.  
673 *Lycopersicon*). *Evol. Int. J. Org. Evol.* 62, 2995–3013. <https://doi.org/10.1111/j.1558-5646.2008.00487.x>
- 674  
675 Moyle, L.C., Nakazato, T., 2010. Hybrid Incompatibility “Snowballs” Between *Solanum*  
676 Species. *Science* 329, 1521–1523. <https://doi.org/10.1126/science.1193063>

- 677 Moyle, L.C., Nakazato, T., 2008. Comparative Genetics of Hybrid Incompatibility: Sterility in  
678 Two Solanum Species Crosses. *Genetics* 179, 1437–1453.  
679 <https://doi.org/10.1534/genetics.107.083618>
- 680 Payseur, B.A., Rieseberg, L.H., 2016. A genomic perspective on hybridization and speciation.  
681 *Mol. Ecol.* 25, 2337–2360. <https://doi.org/10.1111/mec.13557>
- 682 Pease, J.B., Haak, D.C., Hahn, M.W., Moyle, L.C., 2016. Phylogenomics Reveals Three  
683 Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biol.* 14, e1002379.  
684 <https://doi.org/10.1371/journal.pbio.1002379>
- 685 Pease, J.B., Rosenzweig, B.K., 2018. Encoding Data Using Biological Principles: The  
686 Multisample Variant Format for Phylogenomics and Population Genomics. *IEEE/ACM*  
687 *Trans. Comput. Biol. Bioinform.* 15, 1231–1238.  
688 <https://doi.org/10.1109/TCBB.2015.2509997>
- 689 Rick, C.M., 1979. Biosystematic studies in *Lycopersicon* and closely related species of  
690 *Solanum*. *Linn. Soc. Symp. Ser.*
- 691 Rick, C.M., 1950. Pollination Relations of *Lycopersicon Esculentum* in Native and Foreign  
692 Regions. *Evolution* 4, 110–122. <https://doi.org/10.1111/j.1558-5646.1950.tb00046.x>
- 693 Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., Bierne, N., 2016. Shedding Light  
694 on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLOS Biol.*  
695 14, e2000234. <https://doi.org/10.1371/journal.pbio.2000234>
- 696 RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
- 697 Ruhsam, M., Hollingsworth, P.M., Ennos, R.A., 2011. Early evolution in a hybrid swarm  
698 between outcrossing and selfing lineages in *Geum*. *Heredity* 107, 246–255.  
699 <https://doi.org/10.1038/hdy.2011.9>
- 700 Stebbins, G.L., 1970. Variation and Evolution in Plants: Progress During the Past Twenty  
701 Years, in: Hecht, M.K., Steere, W.C. (Eds.), *Essays in Evolution and Genetics in Honor*  
702 *of Theodosius Dobzhansky: A Supplement to Evolutionary Biology*. Springer US,  
703 Boston, MA, pp. 173–208. [https://doi.org/10.1007/978-1-4615-9585-4\\_6](https://doi.org/10.1007/978-1-4615-9585-4_6)
- 704 Suarez-Gonzalez, A., Lexer, C., Cronk, Q.C.B., 2018. Adaptive introgression: a plant  
705 perspective. *Biol. Lett.* 14. <https://doi.org/10.1098/rsbl.2017.0688>
- 706 Taylor, I.B., 1986. Biosystematics of the tomato, in: Atherton, J.G., Rudich, J. (Eds.), *The*  
707 *Tomato Crop: A Scientific Basis for Improvement*, The Tomato Crop. Springer  
708 Netherlands, Dordrecht, pp. 1–34. [https://doi.org/10.1007/978-94-009-3137-4\\_1](https://doi.org/10.1007/978-94-009-3137-4_1)
- 709 The Heliconius Genome Consortium, Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D.,  
710 Davey, J.W., Whibley, A., Nadeau, N.J., Zimin, A.V., Hughes, D.S.T., Ferguson, L.C.,  
711 Martin, S.H., Salazar, C., Lewis, J.J., Adler, S., Ahn, S.-J., Baker, D.A., Baxter, S.W.,  
712 Chamberlain, N.L., Chauhan, R., Counterman, B.A., Dalmay, T., Gilbert, L.E., Gordon,  
713 K., Heckel, D.G., Hines, H.M., Hoff, K.J., Holland, P.W.H., Jacquín-Joly, E., Jiggins,  
714 F.M., Jones, R.T., Kapan, D.D., Kersey, P., Lamas, G., Lawson, D., Mapleson, D.,  
715 Maroja, L.S., Martin, A., Moxon, S., Palmer, W.J., Papa, R., Papanicolaou, A., Pauchet,  
716 Y., Ray, D.A., Rosser, N., Salzberg, S.L., Supple, M.A., SurrIDGE, A., Tenger-Trolander,  
717 A., Vogel, H., Wilkinson, P.A., Wilson, D., Yorke, J.A., Yuan, F., Balmuth, A.L., Eland,  
718 C., Gharbi, K., Thomson, M., Gibbs, R.A., Han, Y., Jayaseelan, J.C., Kovar, C.,  
719 Mathew, T., Muzny, D.M., Onger, F., Pu, L.-L., Qu, J., Thornton, R.L., Worley, K.C.,  
720 Wu, Y.-Q., Linares, M., Blaxter, M.L., French-Constant, R.H., Joron, M., Kronforst, M.R.,  
721 Mullen, S.P., Reed, R.D., Scherer, S.E., Richards, S., Mallet, J., Owen McMillan, W.,



722 Jiggins, C.D., 2012. Butterfly genome reveals promiscuous exchange of mimicry  
723 adaptations among species. *Nature* 487, 94–98. <https://doi.org/10.1038/nature11041>  
724 The Tomato Genome Consortium, 2012. The tomato genome sequence provides insights into  
725 fleshy fruit evolution. *Nature* 485, 635–641. <https://doi.org/10.1038/nature11119>  
726 Turissini, D.A., Matute, D.R., 2017. Fine scale mapping of genomic introgressions within the  
727 *Drosophila yakuba* clade. *PLOS Genet.* 13, e1006971.  
728 <https://doi.org/10.1371/journal.pgen.1006971>  
729 Winger, B.M., 2017. Consequences of divergence and introgression for speciation in Andean  
730 cloud forest birds. *Evol. Int. J. Org. Evol.* 71, 1815–1831.  
731 <https://doi.org/10.1111/evo.13251>  
732  
733

734 Table 1. Introgression statistics for each analyzed trio (4-taxon test), based on data from 100kb windows with > 20  
 735 informative sites. Trios in bold have D statistics that are significantly different than zero after Bonferroni correction. The  
 736 order in which each species accession is listed corresponds to (P1, P2, P3). In all instances, we use the potato genome  
 737 (*S. tuberosum*) as the outgroup.

Group	Accessions	MeanD	StnDev	SE	C95_lwr	C95_upr	Chi-square	p.value
<i>a. Geographic Trios*</i>								
gal.gal.che	LA1044.LA0483.LA0746	-0.0277	0.4947	0.0430	-0.1120	0.0566	0.97	1
<b>arc.arc.pim</b>	<b>LA2172.LA2157.LA2147</b>	<b>0.0328</b>	<b>0.5374</b>	<b>0.0072</b>	<b>0.0186</b>	<b>0.0470</b>	<b>1666.13</b>	<b>2.2e-16</b>
<b>pim.pim.neo</b>	<b>LA1375.LA1246.LA2133</b>	<b>-0.0892</b>	<b>0.5902</b>	<b>0.0073</b>	<b>-0.1036</b>	<b>-0.0749</b>	<b>102.83</b>	<b>5.51E-23</b>
pim.pim.chi	LA1582.LA1933.LA1969	-0.0100	0.5168	0.0094	-0.0285	0.0085	2.82	1
<b>pim.pim.cor1</b>	<b>LA0400.LA1269.LA0118</b>	<b>0.0162</b>	<b>0.5075</b>	<b>0.0080</b>	<b>0.0005</b>	<b>0.0318</b>	<b>10.30</b>	<b>1.82E-02</b>
<b>pim.pim.cor2</b>	<b>LA1617.LA1521.LA0118</b>	<b>0.0998</b>	<b>0.6441</b>	<b>0.0082</b>	<b>0.0838</b>	<b>0.1158</b>	<b>126.44</b>	<b>3.44E-28</b>
<b>pim.pim.per1</b>	<b>LA1595.LA1341.LA1278</b>	<b>0.0047</b>	<b>0.5091</b>	<b>0.0070</b>	<b>-0.0090</b>	<b>0.0184</b>	<b>21.44</b>	<b>5.10E-05</b>
<b>pim.pim.per2</b>	<b>LA1617.LA1269.LA1278</b>	<b>0.0828</b>	<b>0.6145</b>	<b>0.0081</b>	<b>0.0669</b>	<b>0.0987</b>	<b>85.52</b>	<b>3.21E-19</b>
<b>pim.pim.hab</b>	<b>LA0417.LA0442.LA1777</b>	<b>0.0421</b>	<b>0.5683</b>	<b>0.0072</b>	<b>0.0280</b>	<b>0.0561</b>	<b>96.14</b>	<b>1.49E-21</b>
<b>pim.pim.pen</b>	<b>LA1245.LA1269.LA1272</b>	<b>0.0683</b>	<b>0.5357</b>	<b>0.0091</b>	<b>0.0504</b>	<b>0.0862</b>	<b>72.93</b>	<b>1.87E-16</b>
<b>arc.arc.hab</b>	<b>LA2172.LA2157.LA1718</b>	<b>-0.0154</b>	<b>0.4198</b>	<b>0.0074</b>	<b>-0.0299</b>	<b>-0.0009</b>	<b>22.35</b>	<b>3.17E-05</b>
<b>hab.hab.neo</b>	<b>LA1777.LA1718.LA2133</b>	<b>0.0483</b>	<b>0.5144</b>	<b>0.0072</b>	<b>0.0342</b>	<b>0.0624</b>	<b>47.32</b>	<b>8.42E-11</b>
<b>hab.hab.cor</b>	<b>LA0407.LA1777.LA0118</b>	<b>0.0579</b>	<b>0.4796</b>	<b>0.0072</b>	<b>0.0439</b>	<b>0.0720</b>	<b>161.13</b>	<b>8.95E-36</b>
<b>hua.hua.hab</b>	<b>LA1983.LA1365.LA1718</b>	<b>0.0532</b>	<b>0.4348</b>	<b>0.0076</b>	<b>0.0383</b>	<b>0.0680</b>	<b>364.90</b>	<b>3.37E-80</b>
<i>b. Mating System Trios</i>								
<b>arcSl.arcSC.pimSC</b>	<b>LA2172.LA2157.LA0373</b>	<b>0.04356</b>	<b>0.531</b>	<b>0.282</b>	<b>0.007</b>	<b>0.030</b>	<b>1999.06</b>	<b>2.2e-16</b>
<b>arcSl.arcSC.pimSC2</b>	<b>LA2172.LA2157.LA0400</b>	<b>0.0264</b>	<b>0.547</b>	<b>0.008</b>	<b>0.011</b>	<b>0.042</b>	<b>736.04</b>	<b>1.31E-161</b>
<b>habSl.habSC.pimSC</b>	<b>LA1777.LA0407.LA0373</b>	<b>-0.03551</b>	<b>0.508</b>	<b>0.258</b>	<b>0.006</b>	<b>-0.048</b>	<b>42449.7</b>	<b>2.2e-16</b>
<b>habSl.habSC.pimSC2</b>	<b>LA1777.LA0407.LA0400</b>	<b>-0.0357</b>	<b>0.560</b>	<b>0.007</b>	<b>-0.050</b>	<b>-0.021</b>	<b>5526.61</b>	<b>2.2e-16</b>
<b>perSl.perSC.pimSC</b>	<b>LA1278.PI128650.LA0373</b>	<b>0.14860</b>	<b>0.554</b>	<b>0.307</b>	<b>0.008</b>	<b>0.134</b>	<b>5456.4</b>	<b>2.2e-16</b>
<b>perSl.perSC.pimSC2</b>	<b>LA1278.PI128650.LA0400</b>	<b>0.1526</b>	<b>0.562</b>	<b>0.010</b>	<b>0.133</b>	<b>0.173</b>	<b>8.79</b>	<b>9.10E-03</b>

738 StnDev = standard deviation; SE = standard error of mean; C95\_lwr and C95\_upr = Lower and upper 95% confidence intervals,  
 739 respectively, Chi-square = Chi-squared goodness-of-fit test value. \* For geographic trios with non-overlapping accessions, see  
 740 supplemental table 7.

741 FIGURE LEGENDS

742

743 Figure 1. Structured ABBA-BABA tests to evaluate *a priori* hypotheses about the  
744 expected prevalence of introgression due to geographical proximity and/or lineage  
745 differences in mating system. For example, if introgression occurs more frequently  
746 between geographically closer accessions, more minority topologies should support a  
747 closer relationship between P2 and P3, compared to P1 and P3, and the genome-wide  
748 mean D-statistic is expected to be positive.

749

750 Figure 2. Effect of geographical proximity on introgression. For each geographic trio, the  
751 plot shows genome-wide mean *D* values and 95% confidence intervals, as well as *D*-  
752 statistic estimates from individual 100kb windows (gray circles: individual window *D*  
753 values not significantly different than zero; blue circles: individual window *D* values  
754 significantly different than zero). Accessions in each trio are listed in Table 1. For all  
755 trios, our outgroup is potato (*S. tuberosum*).

756

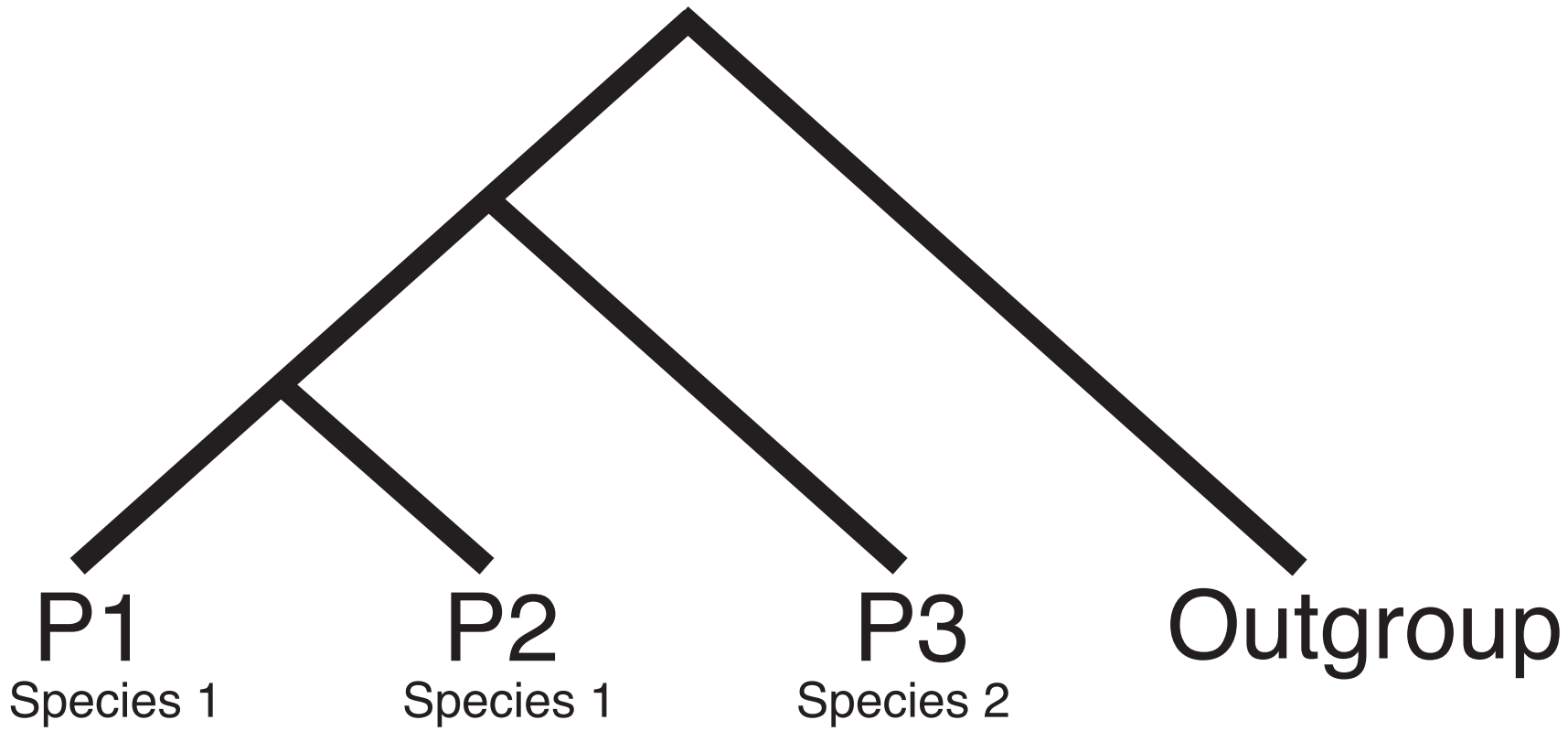
757 Figure 3. Effect of mating system differences on the observed direction of introgression.  
758 For each mating system trio, the plot shows genome-wide mean *D* and 95% confidence  
759 intervals, as well as *D*-statistic estimates from individual 100kb windows (gray circles:  
760 individual window *D* values not significantly different than zero; blue circles: individual  
761 window *D* values significantly different than zero). Here the P3 position is occupied by  
762 the same accession of *S. pimpinifolium* (LA0373) for all three comparisons.

763

764 Figure 4. The relationship between genome-wide mean *D* and the average genetic  
765 distance (% divergence across all sites) between P1/P2 and P3 species for 14 trios  
766 used in geographic tests (R-squared = 0.012, p-value = 0.68).

767





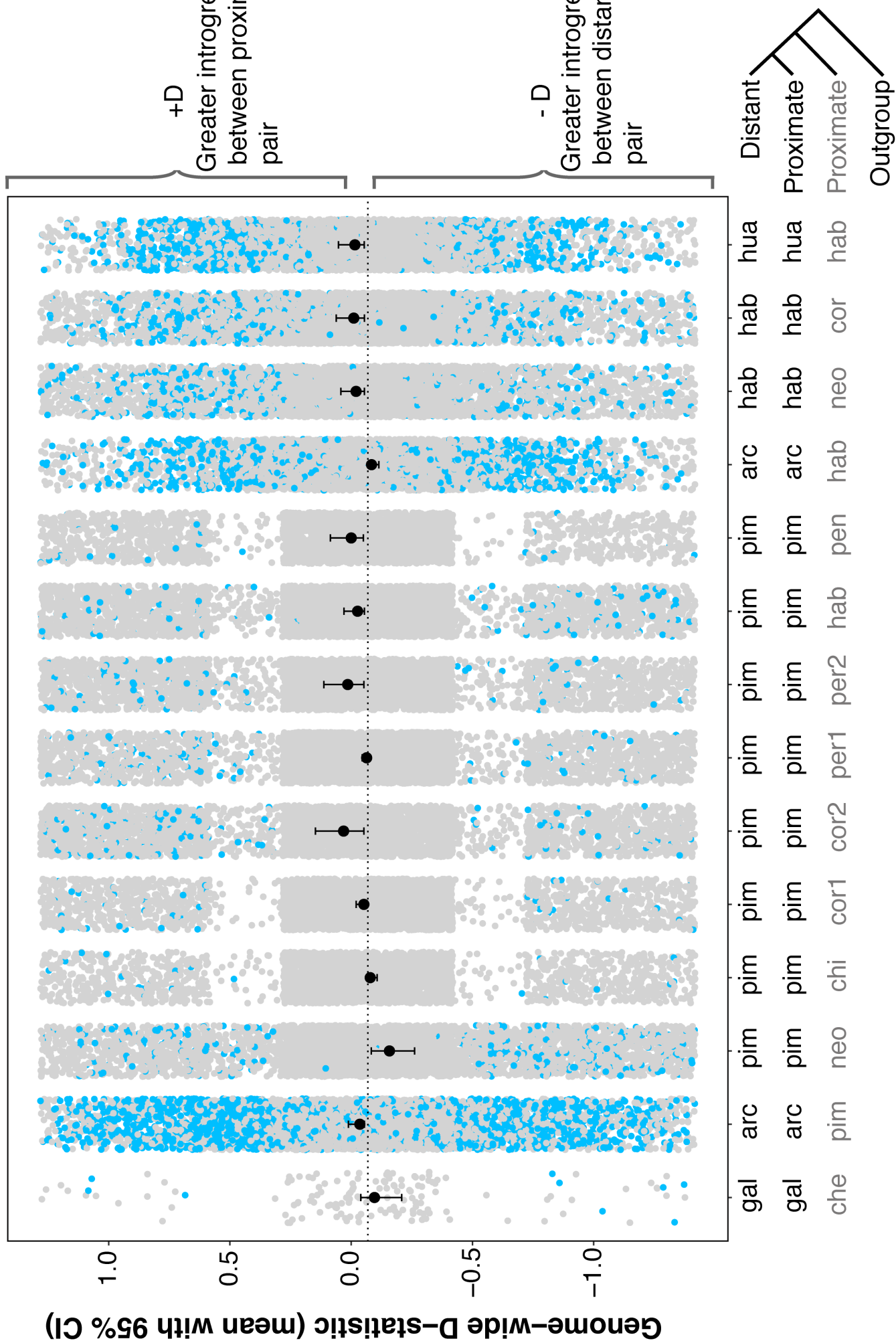
---

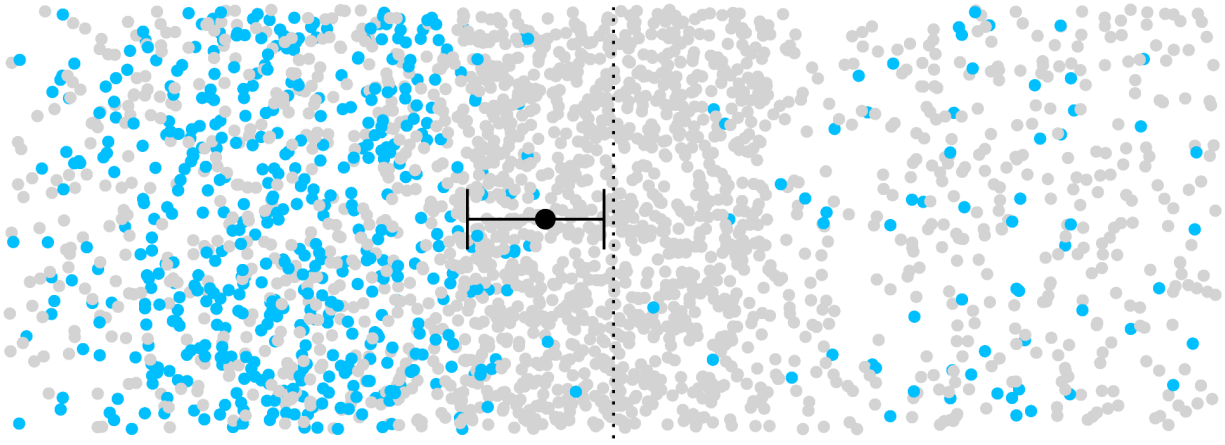
Distant  
SI

Proximate  
SC

Proximate  
SC

Geography  
Mating System

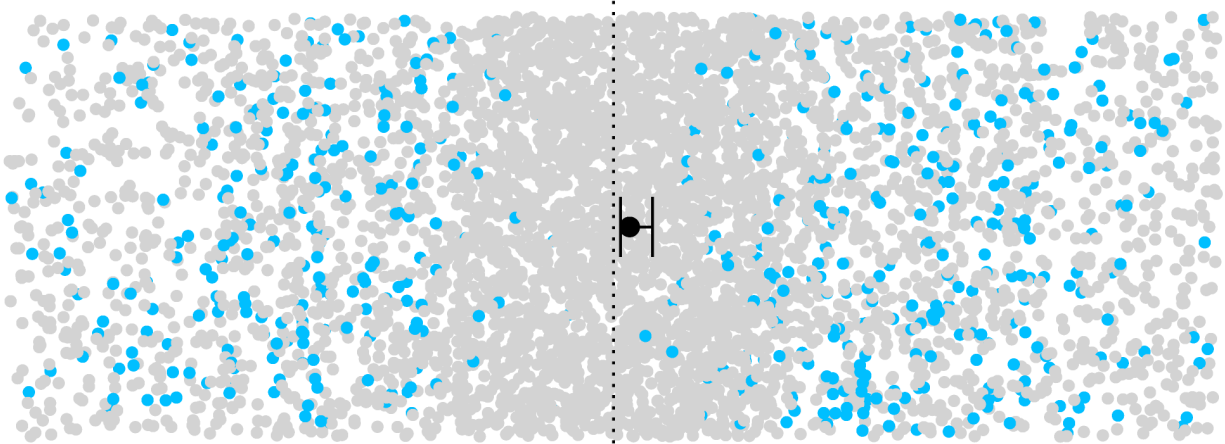




per.SI

per.SI

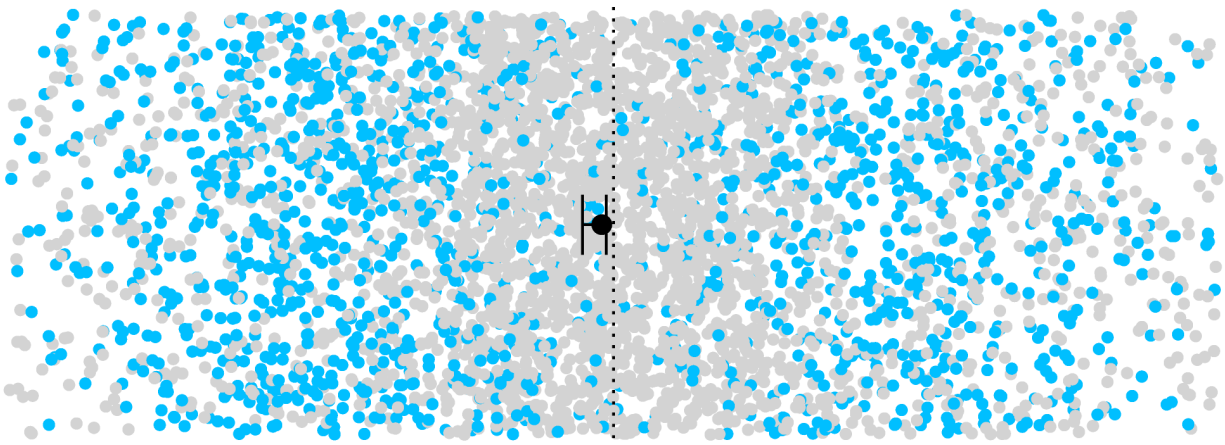
pim.SC



hab.SI

hab.SC

pim.SC



arc.SI

arc.SC

pim.SC

Genome-wide D-statistic (mean with 95% CI)

1.0

0.5

0.0

-0.5

-1.0

