

1           **Model-based species delimitation: are coalescent species reproductively isolated?**

2  
3                   Luke C. Campillo<sup>1\*</sup>, Anthony J. Barley<sup>1</sup>, Robert C. Thomson<sup>1</sup>

4                   <sup>1</sup>*Department of Biology, University of Hawai'i, Honolulu, HI, 96822*

5  
6           \*Corresponding Author: [campillo@hawaii.edu](mailto:campillo@hawaii.edu)

7  
8           **Keywords:** Multispecies Coalescent; BPP; genetic distance; *Drosophila* speciation

9  
10           **ABSTRACT**

11           A large and growing fraction of systematists define species as independently evolving lineages  
12           that may be recognized by analyzing the population genetic history of alleles sampled from  
13           individuals belonging to those species. This has motivated the development of increasingly  
14           sophisticated statistical models rooted in the multispecies coalescent process. Specifically, these  
15           models allow for simultaneous estimation of the number of species present in a sample of  
16           individuals and the phylogenetic history of those species using only DNA sequence data from  
17           independent loci. These methods hold extraordinary promise for increasing the efficiency of  
18           species discovery, but require extensive validation to ensure that they are accurate and precise.  
19           Whether the species identified by these methods correspond to the species that would be  
20           recognized by alternative species recognition criteria (such as measurements of reproductive  
21           isolation) is currently an open question, and a subject of vigorous debate. Here we perform an  
22           empirical test of these methods by making use of a classic model system in the history of  
23           speciation research, flies of the genus *Drosophila*. Specifically, we use the uniquely  
24           comprehensive data on reproductive isolation that is available for this system, along with DNA  
25           sequence data, to ask whether *Drosophila* species inferred under the multispecies coalescent  
26           model correspond to those recognized by many decades of speciation research. We found that  
27           coalescent based and reproductive isolation based methods of inferring species boundaries are  
28           concordant for 77% of the species pairs. We explore and discuss potential explanations for these  
29           discrepancies. We also found that the amount of prezygotic isolation between two species is a  
30           strong predictor of the posterior probability of species boundaries based on DNA sequence data,  
31           regardless of whether the species pairs are sympatrically or allopatrically distributed.

## 32 INTRODUCTION

33 Due to the complex nature of the processes that drive speciation, the development of an  
34 operational species concept is difficult. For many years the Biological Species Concept (Mayr  
35 1963), or some variation of it requiring high (but potentially incomplete) levels of reproductive  
36 isolation between populations, was the standard (Coyne and Orr 2004). Starting in the late 1990s,  
37 work echoing earlier evolutionary species concepts (Simpson 1951; Wiley 1978) led many  
38 evolutionary biologists to recognize that species might best be recognized as “metapopulation  
39 lineages” using any number of data types (de Queiroz 1998, 2007). Under this metapopulation  
40 conceptualization of species, the boundary between independent lineages (i.e., species) is  
41 demarcated by any number of criteria, which may or may not include reproductive isolation.  
42 Despite these conceptual changes, systematists remain in need of an operational framework for  
43 species delimitation. One option for placing the practice of species delimitation in a more  
44 explicit framework is to use statistical models of evolutionary relationships among genetic loci.  
45 One of the most widely used methods for doing so, BPP (Yang and Rannala 2014; Rannala and  
46 Yang 2017), is a Bayesian method based on the multispecies coalescent model (MSC). BPP can  
47 be employed to analyze variation in genealogies among many loci to identify independent  
48 coalescent lineages as species (Leaché and Fujita, 2010; Rannala and Yang, 2003).

49         These methods hold promise for greatly increasing the pace and accuracy of species  
50 description, although their utility depends on whether the lineages identified correspond to  
51 biologically defensible (or meaningful) species *and* the precision of these estimates. A number of  
52 studies have looked at the efficacy of the MSC for species tree inference (e.g., Fujita et al. 2012;  
53 Reid et al. 2014; Barley et al. 2018), but the concordance between independent coalescent  
54 lineages and species boundaries has been less well characterized in the literature. Relevant  
55 studies generally take one of two forms: 1) species boundaries identified under the MSC are

56 compared to those identified by alternative approaches, with concordance viewed as an  
57 indication of accuracy (e.g., Camargo et al. 2012; Willis 2017); or 2) Data simulated under  
58 known demographic scenarios are analyzed under the MSC and the results are used to identify  
59 the relative over- or under-splitting of lineages by the model (Sukumaran and Knowles 2017;  
60 Barley et al. 2018; Luo et al. 2018). Both types of studies frequently find that the MSC identifies  
61 more lineages as species than some consider appropriate (Leaché et al. 2019). This discrepancy,  
62 compounded by the lack of a fully objective species criterion, has led to an increasing wariness  
63 that MSC delimitation may inflate the number of species, often beyond what many systematists  
64 find realistic.

65         These studies have been useful for characterizing the statistical performance of the MSC,  
66 but can be difficult to interpret in the context of real species radiations. On one hand, there is the  
67 possibility that the MSC identified genuine species that were missed by earlier studies because  
68 they are morphologically cryptic or otherwise difficult to differentiate. On the other hand, in  
69 simulation studies, the accuracy of results relies on the generating model (i.e., the simulation)  
70 being an accurate representation of biological reality. If important aspects of species biology are  
71 missing from the generating model we may draw inaccurate conclusions merely because our  
72 simulation study is overly simple relative to nature. The challenges associated with both of these  
73 approaches might be averted by turning to an empirical system where extensive data on the  
74 nature of species boundaries themselves are available. In such a system, we can examine the  
75 correspondence between lineages identified as being independent under the MSC and other  
76 criteria placed on species boundaries, such as reproductive isolation.

77         Here we make use of an empirical study system in which an extraordinarily rich  
78 understanding of reproductive isolation has been established for many species pairs, the genus  
79 *Drosophila*. We use this to compare species identification based on reproductive isolation with

80 species delimitation under the MSC. *Drosophila* is an intensively studied model organism in  
81 many disciplines, and arguably the most influential clade for understanding the consequences of  
82 reproductive isolation as a mechanism of species formation and perpetuation. Much of this focus  
83 stems from the fact that experimental lab crosses are possible between species, which provides  
84 information on which species pairs show intrinsic reproductive isolation (lab crosses do not take  
85 extrinsic factors into account) and to what degree that isolation is complete. Two classic papers  
86 by Coyne and Orr (1989, 1997) synthesized data on reproductive isolation derived from  
87 experimental crosses and allozyme electrophoretic distance, which they used as a proxy for time  
88 since divergence by assuming a constant molecular clock. These data were used to produce  
89 estimates for the amount of total reproductive isolation (the combined accumulation of pre- and  
90 postzygotic isolation) and time (using genetic distance as a proxy) required for complete  
91 speciation. Here, complete speciation refers to a relaxed version of Mayr's Biological Species  
92 Concept (Mayr 1963) that allows for low levels of gene flow between closely related species (see  
93 also Coyne and Orr 2004).

94         These studies provide insight into the nature of how species boundaries themselves  
95 become established. A central finding from Coyne and Orr (1989, 1997), and the one we focus  
96 on here, is that the amount of reproductive isolation increases with time since divergence.  
97 Building off this idea, we examined several aspects of how species identified under the MSC  
98 model matched those identified based on experimental quantification of reproductive isolation  
99 and genetic distance. Specifically, we were interested in measuring the correlation between the  
100 amount of reproductive isolation and genetic distance between two species, and the posterior  
101 probability that those two species were identified as independent coalescent lineages. By  
102 quantifying the relationship between reproductive isolation and coalescent lineages we stand to  
103 gain a better understanding of the relationship between the biological processes that drive

104 speciation and promising statistical approaches to species delimitation. It should also be noted  
105 that most studies of reproductive isolation look at comparisons between species, whereas most  
106 practitioners of MSC species delimitation think about comparisons within species (i.e., at the  
107 population level). This study is explicitly concerned with delimitation of nominal species.

108         Assessing the accuracy of species delimitation remains difficult because speciation is a  
109 continuous process and species boundaries are not always discrete. However, utilizing  
110 information on reproductive isolation allows us to assess the performance of these methods in a  
111 more empirically grounded framework than is otherwise possible. With this in mind, the research  
112 goals for this study were to: 1) examine the relationship between coalescent and reproductive  
113 isolation based species delimitation; 2) quantify how varying levels of pre- and postzygotic  
114 isolation in allopatric and sympatric species pairs affect our ability to identify species under the  
115 MSC; 3) better understand the utility of MSC methods to increase the accuracy and precision of  
116 species recognition in the presence of partial, or complete, reproductive isolation.

117

## 118 **MATERIALS & METHODS**

### 119 **Dataset Assembly**

120 The lists of species pairs studied by Coyne and Orr (1989, 1997) were combined to assemble a  
121 set to be used in the present study ( $n = 108$  pairwise comparisons). Values for pre- and  
122 postzygotic isolation were compiled from Coyne and Orr's two studies (1989, 1997), and  
123 updated based on isolation measures from a more recent study (Yukilevich 2012). Total  
124 reproductive isolation ( $T$ ; Coyne and Orr 1989) was calculated using the equation:  $T = Pre +$   
125  $(1 - Pre) \times Post$ . In keeping with previous work, if a species pair only had data on prezygotic  
126 or postzygotic isolation, and that value was greater than 0.95 or 1.0, respectively, we considered  
127  $T$  to equal the measure for which there was data (Coyne and Orr 1997). Conversely, if a species

128 pair only had data for prezygotic or postzygotic isolation, and that value was less than 0.95 or  
129 1.0, respectively, those species pairs were excluded from downstream analysis. Data on allozyme  
130 genetic distance (D) comes originally from Coyne and Orr (1989, 1997), but more recently  
131 updated by Yukilevich (2012).

132 For the purpose of this study, we follow the species recognition thresholds proposed by  
133 Coyne and Orr (1997). Specifically, they proposed minimum values on total reproductive  
134 isolation ( $T \geq 0.903$ ) and genetic distance (sympatric pairs:  $D \geq 0.04$ ; allopatric pairs:  $D \geq 0.54$ )  
135 required for maintenance of species boundaries. While we recognize that the use of any  
136 particular threshold may raise concerns, we opted to use these for two reasons. First, by using the  
137 same metrics employed by previous studies, we are able to make a direct comparison between  
138 species delimitation based on reproductive isolation and delimitation based on MSC models.  
139 Second, because MSC species delimitation has many practical implications, we wanted to focus  
140 on values that would likely be the most relevant and widely used in empirical studies

141 We assembled a set of genes ( $n=8$ ) that have been sequenced across a wide taxonomic  
142 breadth within *Drosophila* (van der Linde and Houle 2008; Yang et al. 2012). We then  
143 downloaded whatever sequence data was available from GenBank for this set of genes for all  
144 species ( $n = 59$ ) on the compiled list. Custom Python and R scripts (R Core Team 2013) were  
145 used for data cleanup. Species pairs were then sorted into eleven species groups for analysis  
146 (Coyne and Orr 1997). Here, “species group” refers to a term used in *Drosophila* literature to  
147 reference a monophyletic set of closely-related species (see O’Grady and DeSalle 2018). While a  
148 large amount of mitochondrial data is available for *Drosophila*, we chose to include only one  
149 mitochondrial gene (Cytochrome c oxidase subunit II: COII) for each species group because we  
150 expect the entire mitochondrial genome to share a single coalescent history. We limited the final  
151 data matrix for each group to include only those genes that were missing sequence data for no

152 more than two species. Because some intensively studied species have an abundance of sequence  
153 data relative to the other species in their group (e.g., *Drosophila subobscura* has 137 COII  
154 sequences available, while the other six species in the Obscura group combine for 30 available  
155 sequences), we pruned sequences from overrepresented taxa to make the amount of data more  
156 even across species. Specifically, for species that had >10 sequences at a given locus, we  
157 randomly pruned sequences until the number that remained was equal to the next most densely  
158 sampled species. Sequence data was aligned for each group using MUSCLE (Edgar 2004) and  
159 converted to PHYLIP format using DendroPy 4.2.0 (Sukumaran and Holder 2010). These eleven  
160 matrices, one for each species group, were analyzed independently in all downstream analyses.  
161 To check for any difference in electrophoretic distance and genetic distances calculated from  
162 DNA sequence data, we calculated Jukes-Cantor genetic distances across all species pairs for  
163 COII in Geneious v 7.1.9 (Kearse et al. 2012) and then compared those to Nei's D (Fig. S1;  
164 Table S1).

165

## 166 **MSC Analysis**

167 After assembling the dataset, we performed MSC species delimitation using BPP v 3.2 (Yang  
168 and Rannala 2014) on each of the 11 species groups separately. BPP uses reversible-jump  
169 Markov Chain Monte Carlo (rjMCMC) to estimate a Bayesian posterior distribution for different  
170 species delimitation models. We jointly estimated the species tree topology and assignment of  
171 individuals into species (referred to as “analysis A11” in BPP). Under this model we define a  
172 starting tree, where each named species in the group is a “population”, and then use the rjMCMC  
173 to target the posterior distribution of possible models that may merge combinations of these  
174 populations into a smaller number of species. This approach does not allow single populations to  
175 be split more finely, and so the existing taxonomy within each species group (i.e., named species

176 from Coyne and Orr 1997) forms the upper limit on the total number of possible species in each  
177 analysis. We specified a prior distribution that assigns equal probability for the number of  
178 species (i.e., all populations merged into one species or a different species for each population)  
179 and then divides that probability by the proportion of compatible labeled histories  
180 (*speciesmodelprior* = 2; Yang 2015). We also set a prior that allows  $\theta$  (population size  
181 parameter) to vary among loci according to specified heredity multipliers (Hey and Nielsen  
182 2004) to account for the differences in effective population size between nuclear and  
183 mitochondrial loci (*heredity* = 2; Yang and Rannala 2014).

184 Two central parameters of the MSC model are for population sizes ( $\theta$ ) and species  
185 divergence times ( $\tau$ ). These are specified as gamma distributed random variables in BPP. Our  
186 goal was to set up a diffuse, but credible, prior distribution for both of these parameters. Here,  
187 we constructed an empirical prior for  $\theta$  in *Drosophila* by using published estimates of this value,  
188 if available, or published estimates of effective population size ( $N_e$ ) and mutation rate ( $\mu$ ) to  
189 solve for the equation:  $\theta = 4N_e\mu$  (Eyre-Walker et al. 2002; Wall et al. 2002; Tamura 2003; Yi et  
190 al. 2003; Hey and Nielsen 2004; Haag-Liautard et al. 2007; Pascual et al. 2007; Cutter 2008;  
191 Charlesworth 2009; Keightley et al. 2009, 2014; Legrand et al. 2009; Obbard et al. 2012; Smith  
192 et al. 2012; Schrider et al. 2013). Similarly, to obtain an empirical prior for the  $\tau$  parameter, we  
193 used the online database TimeTree ([www.timetree.org](http://www.timetree.org)) to obtain an estimated root age for each  
194 species group. Because each species group had different estimated root ages, there was a unique  
195 combination of  $\theta$  and  $\tau$  prior gamma distributions for each group. For both prior distributions,  
196 we centered the mean of the gamma distribution on the empirical values, and set the variance of  
197 the distribution to be wider than the distribution of all empirical values. We ran 204,000 MCMC  
198 generations, discarding the first 4000 generations as burnin and then sampling every 20  
199 generations until we reached 10,000 samples (BPP output files will be available on Dryad Digital



200 Repository upon publication). We repeated each analysis two times, checking that the results  
201 remained consistent across runs (Yang 2015).

202         The BPP manual (Yang 2015) explicitly states that there are no default priors for  $\theta$  and  $\tau$ ,  
203 yet many studies employ a set of three prior gamma distributions originally used by Leaché and  
204 Fujita (2010) as if they were. Here we explored prior sensitivity by performing additional BPP  
205 runs on each dataset under this suite of priors (Leaché and Fujita 2010). For the sake of clarity,  
206 we will refer to the set of priors from Leaché and Fujita (2010) as empirically “uninformed” and  
207 our empirically derived priors as “informed”. In total, we ran a BPP delimitation analysis on  
208 each of the eleven species groups under four unique prior settings (one informed, three  
209 uninformed) resulting in 44 independent BPP runs. Within the uninformed set of priors, we focus  
210 on those that most closely matched our prior expectations about the coalescent history for  
211 *Drosophila* (large ancestral population and shallow divergence among populations in this case).  
212 This should minimize differences between informed and uninformed priors, thereby supplying a  
213 realistic (and somewhat conservative) test of prior sensitivity. Full details for all prior settings  
214 can be found in supplementary material (Table S2), including  $\theta$  and  $\tau$  priors for all groups and  
215 BPP runs. We extracted posterior  $\theta$  values from each MCMC output file using a custom Python  
216 script. To summarize the posterior distribution for  $\tau$ , we used the R package Phytools v 0.6-60  
217 (Revell 2012) to read the sampled trees in the MCMC output into R and calculate the tree height  
218 of each (Python and R codes will be available on Dryad Digital Repository upon publication).  
219 We visually inspected concordance between prior and posterior distributions of  $\theta$  and  $\tau$  for the  
220 informed and uninformed priors. Additionally, we investigated whether the number of species  
221 delimited was sensitive to the alternative priors.

222         The A11 analysis in BPP estimates posterior probabilities for different delimitation  
223 models that may differ in number of species and the topology of the species phylogeny. Here, we

224 are specifically interested in the marginal posterior probability for the splitting, or lumping, of  
225 each particular species pair, rather than the marginal posterior probability of any one delimitation  
226 model. Because species pairs can be split or lumped in different configurations, we calculated the  
227 marginal posterior probability of independence between each species in a given pair. We refer to  
228 this value as the Posterior Probability of Independent Lineages (PPIL). PPIL values are  
229 practically, and philosophically, the same measure as the “speciation probability” described in  
230 Leaché and Fujita (2010). The difference is that we calculate these values from analyses that  
231 marginalize across both species delimitations and species tree topologies, whereas Leaché and  
232 Fujita (2010) only marginalized across species delimitations due to limitations of the software at  
233 that time which required the species tree be held constant (i.e., we do not employ a guide tree in  
234 our analysis, as was done previously).

235 We calculated PPILs for all species pairs from Coyne and Orr (1989, 1997) where  
236 sufficient sequence data was available to conduct the analysis, and for which information on  
237 reproductive isolation was available (n = 108 pairwise comparisons). Using PPIL values, we  
238 compared the number of independent lineages identified by the MSC to those based on  
239 reproductive isolation and genetic distance documented by Coyne and Orr (1989, 1997). The two  
240 species in each pair were considered independent lineages if they had a high PPIL ( $\geq 0.95$ ) or  
241 met the criteria for total isolation ( $T \geq 0.903$ ) and genetic distance ( $D_{\text{allo}} \geq 0.54$ ;  $D_{\text{symp}} \geq 0.04$ )  
242 from Coyne and Orr (1997).

243

#### 244 **Reproductive Isolation and Genetic Distance**

245 Besides a cursory comparison in numbers of species delimited by either method, we were also  
246 interested in how the components of speciation considered by Coyne and Orr (1989, 1997)  
247 related to PPIL values in our present study. To accomplish this, we selected among a series of

248 generalized additive models (GAMs) to assess the impact of different predictor variables  
249 (prezygotic isolation, postzygotic isolation, and genetic distance) on the PPIL for allopatric  
250 and sympatrically distributed species pairs. We chose not to include total reproductive isolation  
251 in these models because it is a function of prezygotic and postzygotic isolation, which we already  
252 account for independently. We elected to use a GAM because the model can flexibly capture the  
253 impact of a predictor variable through a smoother function, allowing for both linear and non-  
254 linear relationships. Because the dependent variable (i.e., PPIL) included 0 and 1, we performed  
255 a logit transformation to normalize the data using the R package *car* (Fox and Weisberg 2011).  
256 We then used the R package *mgvc* to construct and fit GAMs to the data (Wood 2011). We  
257 calculated Akaike Information Criterion (Akaike 1974) values for each model and considered the  
258 model with the smallest AIC to be the best fitting model. We further considered models with  
259  $\Delta AIC \leq 2$  from the best model to be similarly plausible (Burnham et al. 2002).

260 We analyzed two independent sets of GAMs to: 1) explore the differences between  
261 allopatric and sympatric species pairs and 2) investigate the impact of predictor variables within  
262 allopatric and sympatric species pairs separately. The first series of GAMs compared models  
263 with one smoother function for all species pairs to models specifying separate smoothers for  
264 allopatric and sympatric species pairs. Improved fit for the latter model would indicate that range  
265 overlap between taxa differentially predicts PPIL values, and therefore these two types of species  
266 pairs (allopatric and sympatric) should be modelled separately in downstream analyses.

267 After confirming two smoothers based on geography was a better model, we fit the  
268 second series of GAMs on allopatric, sympatric, and closely related ( $D \leq 0.50$ ) allopatric and  
269 sympatric species pairs. For each set of species pairs, we fit a total of five GAMs, modelling the  
270 three predictor variables listed previously (prezygotic isolation, postzygotic isolation, and genetic  
271 distance) and two additive models (prezygotic isolation + genetic distance and postzygotic

272 isolation + genetic distance). We fit the GAMs on taxa with low genetic distance ( $D \leq 0.5$ ) to  
273 maintain consistency with Coyne and Orr (1989), and because they suggested closely related  
274 sympatric species pairs should show the strongest signature of reinforcement. In this context,  
275 evidence for reinforcement's effect would be that the model for prezygotic isolation is the  
276 strongest predictor of PPIL among closely related sympatric species, followed closely by a  
277 model of postzygotic isolation. Specifically, because reinforcement acts to increase prezygotic  
278 isolation in order to counteract disadvantageous hybridization, there must be some level of  
279 postzygotic isolation.

280       Because data from species pairs are phylogenetically non-independent, we also explored  
281 how this might impact our results. To do so, we extracted the maximum *a posteriori* tree from  
282 each of the BPP analyses using phytools v 0.6-60 (Revell 2012), setting branch lengths to their  
283 marginal posterior mean. We then assembled a reduced set of phylogenetically independent  
284 species pairs (i.e., only including pairs that are connected by paths on the tree that do not  
285 intersect with other pairs; see Felsenstein 2004, pg. 444). We then reran all the GAM analyses on  
286 this reduced but phylogenetically independent dataset.

287       Lastly, we investigated how PPIL values may vary with respect to levels of pre- and  
288 postzygotic isolation. To do this, we first constructed a reduced dataset that included all species  
289 pairs with prezygotic ( $n = 100$ ) or postzygotic ( $n = 74$ ) isolation data. As outlined by Coyne and  
290 Orr (1989, 1997), both of these values range from 0-1.0, with complete reproductive isolation  
291 (pre- or postzygotic) equal to 1.0. We then performed an ANOVA to determine if PPIL differed  
292 according to the levels of pre- and postzygotic isolation. Because the measures of postzygotic  
293 isolation increased from 0 to 1.0 in 0.25 increments, we treated each level as a separate category  
294 for the ANOVA. For prezygotic isolation, we binned the values based on the same levels (e.g., 0-  
295 0.25, 0.25-0.50, etc.). Coyne and Orr (1989, 1997) investigate what role Haldane's Rule

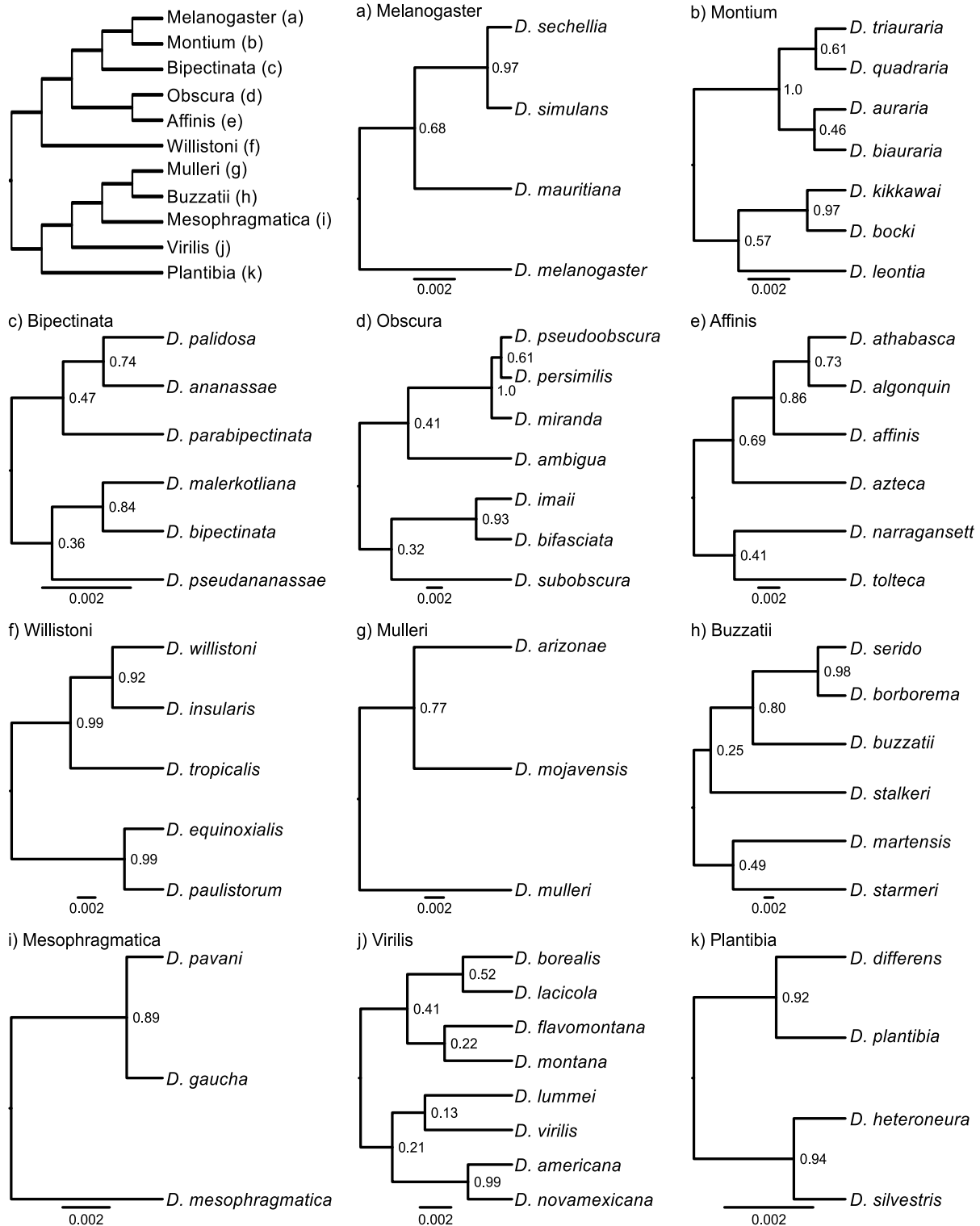
296 (Haldane 1922) may play in the speciation process and find it to be an important early step. We  
297 checked whether it might also influence PPIL by assembling a reduced dataset containing only  
298 those species pairs with 0.25-0.75 postzygotic isolation ( $n = 32$ ), because at these levels of  
299 postzygotic isolation the pair could either be in line with Haldane's Rule (i.e., the heterogametic  
300 sex, males in this case, are either sterile or infertile for both crosses) or not (i.e., a male and  
301 female, or both females are sterile/inviable). Information on sterility/infertility was taken from  
302 Yukilevich (2012). We manually categorized species pairs as either conforming to Haldane's  
303 Rule or not and performed a one-way t-test to look for differences between both categories of  
304 species pairs.

305

## 306 **RESULTS**

### 307 **Genetic dataset assembly and analysis**

308 The final dataset included a total of 543 sequences from 8 genes. We were able to obtain  
309 sequence data for all species from one gene, COII, totaling 160 sequences for that gene. The  
310 median number of genes per species group was three. The Bipectinata group had the greatest  
311 gene diversity ( $n = 6$ ), while the Buzzatii group was represented by only two genes. The mean  
312 number of sequences compared between each species pair in each species group was 7.25,  
313 ranging from an average of three sequences per pair in the Affinis group, up to 13.7 sequences  
314 per pair in the Melanogaster group. Complete information on number of genes per group and loci  
315 per species pair can be found in Supplementary Material (Table S3). The average genetic  
316 distance across all species pairs was 0.82, ranging from 0.026 (*D. heteroneura* – *D. silvestris*) to  
317 1.95 (recorded for multiple species pairs in the Obscura group).



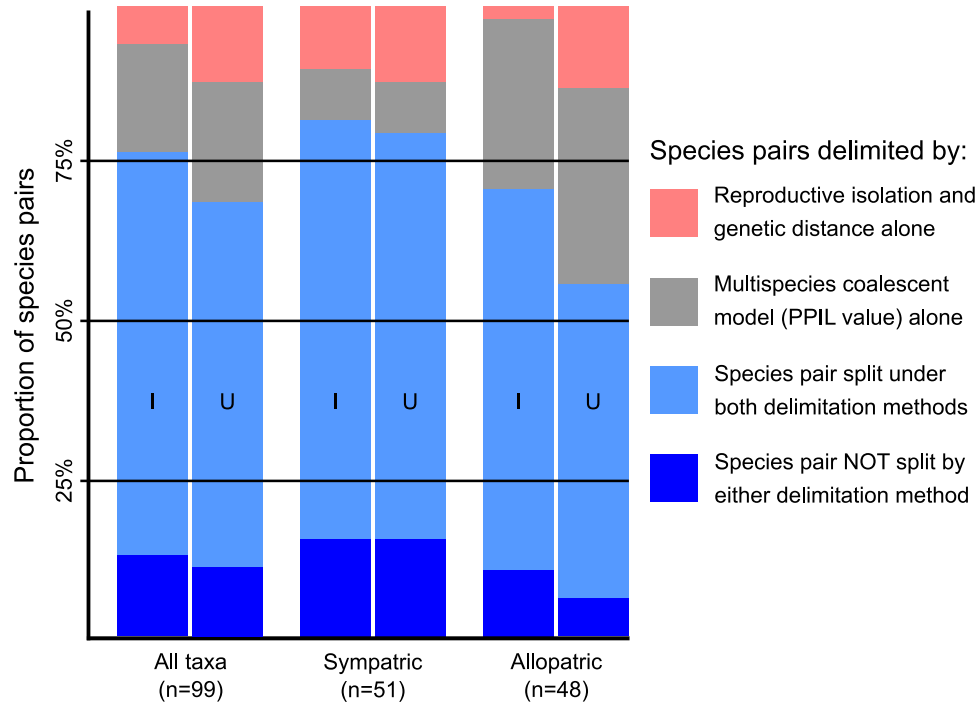
318  
 319 Figure 1) Consensus species trees for each species group in this study. Branch lengths (in coalescent units) are based  
 320 on the average branch length from every MCMC tree that was topologically concordant with the highest probability  
 321 species tree (taken from BPP output). Note that the scale bar is a different length for each tree because it represents  
 322 the same value (0.002 coalescent units) for all trees. Backbone tree (top left) modified from Castillo (2017).

323 We ran the BPP analyses two times for each species group, and obtained similar results  
324 across runs. Consistency across runs indicates that mixing is likely adequate and the MCMC  
325 chain did not get stuck on any one model (Yang 2015). Species trees for each species group  
326 revealed varying levels of posterior probability support across nodes, but overall the species trees  
327 correspond with results from other phylogenetic studies (Fig. 1; van der Linde and Houle 2008;  
328 Yang et al. 2012).

329

### 330 **Concordance of reproductively isolated species and coalescent species delimitation**

331 The two approaches (i.e., threshold of genetic distance and reproductive isolation vs. PPIL  
332 threshold) recognized 63 out of the 99 species pairs as comprising two independent lineages. An  
333 additional 13 species pairs were not recognized as independent lineages by either approach. The  
334 two approaches were therefore in agreement on the species status for 76 of 99 species pairs in  
335 total (~77%; Fig. 2). Of the remaining 23 species pairs, 17 met the threshold for independent  
336 lineages based on PPIL only, whereas the other six species pairs were considered to be distinct  
337 species based on total reproductive isolation (T) and genetic distance (D) only. Qualitatively  
338 similar, but somewhat more discordant, results were recovered when using the uninformed priors  
339 (68% agreement between methods). There was greater concordance between methods when  
340 delimiting sympatric taxa than allopatric taxa (Fig. 2) and this pattern held true across the  
341 informed and uninformed priors. When lowering the admittedly arbitrary, but common, PPIL  
342 threshold from 0.95 to 0.85 we recovered roughly the same amount of concordance (75 out of 99  
343 species pairs in agreement; ~76% concordance). However, there were more lineages recognized  
344 based on PPIL alone (Fig. S2). We also found greater discordance between methods for closely  
345 related species pairs ( $D \leq 0.5$ ) than those more distantly related ( $D > 0.5$ ; Fig. S3).



346  
 347 Figure 2) Plots comparing concordance between species delimitation models. Species delimitation models were in  
 348 agreement for the majority of species pairs (Both; blue). However, there was discrepancy between certain species  
 349 pairs. Some were split solely on reproductive isolation and genetic distance (RI + D; red) while others only on the  
 350 basis of PPIL value  $\geq 0.95$  (MSC; gray). Results are shown for both informed (I) and uninformed (U) prior settings.  
 351

### 352 Reproductive Isolation and Genetic Distance

353 We recovered better predictive models of PPIL when considering sympatric and allopatric taxa  
 354 separately for prezygotic isolation and genetic distance, and similarly plausible models for  
 355 postzygotic isolation (Table 1). When fitting the second set of GAMs to the different types of  
 356 species pairs we found that the models of prezygotic isolation and genetic distance generally  
 357 explained more null deviance than those for postzygotic isolation, which essentially reduced to a  
 358 linear model (Fig. 3; plots for other GAM models are provided in Fig. S4 and Fig. S5). For  
 359 allopatric taxa, the combined effect of prezygotic isolation + genetic distance was the best  
 360 predictor of PPIL, but the effect of prezygotic isolation alone was similarly plausible ( $\Delta AIC =$   
 361 1.11; Table 2). In sympatric taxa, prezygotic isolation was the best predictor of PPIL, followed  
 362 closely by the similarly plausible combination model of prezygotic isolation + genetic distance  
 363 ( $\Delta AIC = 0.21$ ; Table 2). In the sympatric species pairs with low genetic distance ( $D < 0.5$ ), we



364 found that the best predictor of PPIL was again prezygotic isolation, with models of genetic  
 365 distance and postzygotic isolation + genetic distance similarly plausible ( $\Delta AIC = 0.66$  &  $\Delta AIC$   
 366  $= 0.58$ , respectively). We consistently found that the null deviance explained was low for the  
 367 postzygotic isolation models (Table 2). These results remained qualitatively unchanged for the  
 368 reduced phylogenetically independent datasets, with prezygotic isolation being the best predictor  
 369 of PPIL in allopatry and sympatry. We present the results from the full dataset here, while  
 370 complete results from the phylogenetically independent dataset can be found in the  
 371 Supplementary Material (Table S4).

372 When looking at species pairs with prezygotic isolation data ( $n=100$ ), we recovered a  
 373 significant difference between the level of prezygotic isolation and PPIL value, particularly  
 374 between species pairs with lower levels of prezygotic isolation (ANOVA; F-stat = 15.79; p-value  
 375  $= 2.01 \times 10^{-8}$ ; Fig. 4a). However, when considering all species pairs with postzygotic isolation  
 376 data ( $n=74$ ), we found that there was no statistical difference in PPIL among the five levels of  
 377 isolation (ANOVA; F-stat = 1.144; p-value = 0.343; Fig. 4b). Additionally, although the mean  
 378 PPIL for species pairs that follow Haldane’s Rule was lower (0.927,  $n = 21$ ) than those that do  
 379 not (0.97;  $n = 11$ ), there was no statistically significant difference in PPIL between the groups (t-  
 380 test; p-value = 0.179; Fig. S6).

381 **Table 1)**

Number of Smoothers	Prez. Isolation (n = 100)		Post. Isolation (n = 74)		Gen. Distance (n = 108)	
	$\Delta AIC$	Dev. Exp.	$\Delta AIC$	Dev. Exp.	$\Delta AIC$	Dev. Exp.
One (Allo/Symp Together)	8.15	0.29	***	0.07	1.85 <sup>†</sup>	0.22
Two (Allo/Symp Separate)	***	0.40	1.94 <sup>†</sup>	0.07	***	0.27

382  
 383 Table 1) Results from a series of GAMs to test if relationships are different between allopatric and sympatric species  
 384 pairs, where \*\*\* means that model had the lowest AIC value and is therefore the most plausible model. The dagger  
 385 (†) indicates a similarly plausible model ( $\Delta AIC \leq 2$ ; Burnham et al. 2002). The proportion of null deviance  
 386 explained by the model (“Dev. Exp.”) is also provided.

## 387 Effect of the Priors

388 We did not see an effect of the informed versus uninformed priors on the mean number of  
 389 lineages delimited in a given group (i.e., BPP delimited a consistent number of species in each  
 390 group across different prior settings). However, we did see a difference in the variance (i.e., the  
 391 prior influenced how much uncertainty there was in the number of species in each group; Fig.  
 392 5b). Specifically, under the informed priors the MCMC samples were spread more evenly across  
 393 possible delimitation models. We also observed a relatively large difference between the prior  
 394 and posterior distributions of  $\theta$  and  $\tau$  for analysis under uninformed priors, which was not  
 395 observed in analyses under the informed prior (Fig. 5a). This indicates that the uninformed priors  
 396 are a relatively poor match for *Drosophila*, which has the effect of biasing the results to be  
 397 excessively certain. Moreover, this finding highlights that appropriate priors are important for  
 398 accurately estimating the variance of a random variable, in addition to accurately estimating its  
 399 mean.

400 **Table 2)**

Explanatory Variable	Allopatric (n = 28)		Allopatric (D≤0.50; n = 15)		Sympatric (n = 42)		Sympatric (D≤0.50; n = 23)	
	ΔAIC	Dev. Exp.	ΔAIC	Dev. Exp.	ΔAIC	Dev. Exp.	ΔAIC	Dev. Exp.
Prezygotic Isolation	1.11 <sup>†</sup>	0.51	***	0.61	***	0.21	***	0.39
Postzygotic Isolation	14.80	0.08	8.22	0.07	1.25 <sup>†</sup>	0.07	4.86	0.07
Genetic Distance	11.19	0.29	5.34	0.34	2.44	0.05	0.66 <sup>†</sup>	0.33
Pre. Isol. + Gen. Dist.	***	0.57	4.52	0.40	0.21 <sup>†</sup>	0.23	2.23	0.33
Post. Isol. + Gen. Dist.	13.08	0.29	6.18	0.40	2.88	0.08	0.58 <sup>†</sup>	0.39

401  
 402 Table 2) Table of ΔAIC values, where \*\*\* indicates the most plausible model (i.e., which explanatory variable best  
 403 explains PPIL value). The dagger (†) indicates a similarly plausible model (ΔAIC ≤ 2; Burnham et al. 2002). The  
 404 plus sign (+) between variables indicates a model of the additive effect of two predictors. The proportion of null  
 405 deviance explained by the model (“Dev. Exp.”) is also provided.

## 406 DISCUSSION

### 408 Concordance of Species Boundaries Under the MSC

409 We assessed concordance between species that are delimited under population genetic models of  
 410 lineage divergence with a biologically important and widely used measure of species

411 distinctiveness, reproductive isolation. Specifically, we looked at species delimitations inferred  
412 under the BPP implementation of the MSC (which we consider an operational framework for a  
413 lineage-based species concept; Simpson 1951; Wiley 1978; De Queiroz 2007), and those  
414 considered distinct species determined by the amount of reproductive isolation and genetic  
415 distance between two lineages (i.e., a modified Biological Species Concept; Mayr 1963; Coyne  
416 and Orr 2004).

417         We find that the methods do not return identical results, and the multispecies coalescent  
418 tends to more readily split species than measures of reproductive isolation and genetic distance  
419 alone. This could result from the MSC recognizing genuine species that might be missed by  
420 other approaches. Specifically, our results confirm that (if we take MSC delimitation as truth)  
421 reproductive isolation does not need to be complete in order for lineages to be identifiable as  
422 independent. Alternatively, recent studies have found the MSC may be prone to oversplitting and  
423 might result in delimiting “population structure, not species” (Sukumaran and Knowles 2017;  
424 Chambers and Hillis 2019). However, because we are attempting to delimit nominal species any  
425 population structure within species should not mislead our interpretations of the species  
426 delimitation models. That being said, this also means that our results may provide a somewhat  
427 optimistic view of MSC performance.

428         When using the  $PPIL \geq 0.95$  cutoff, both criteria agreed with one another in the majority,  
429 but far from all, cases (76% agreement; PPIL values for all species pairs found in Table S5). Of  
430 these species pairs, 63 were delimited by both methods and 13 were delimited by neither. The  
431 cases where neither delimitation model recognizes these nominal taxa as independent lineages  
432 represent scenarios in which the taxonomy may require reexamination. For example, none of the  
433 species pairs in the *Auraria* species complex were recognized as independent by either method,  
434 suggesting current taxonomy in this species complex may be recognizing too many species.

435 Watada et al. (2011) recently took one step in this direction, revising the taxonomy of *D.*  
436 *quadraria* and suggesting it is better categorized as a junior synonym of *D. triauraria*.

437         Of the remaining species pairs ( $n = 23$ ) for which there is disagreement between methods,  
438 the majority ( $n=17$ ) were pairs split under the MSC (i.e., high PPIL) but not under the Coyne and  
439 Orr criteria. Most of these high PPIL pairs did not have enough total reproductive isolation ( $n =$   
440 14) to meet the Coyne and Orr threshold, while the remaining three species pairs did not have  
441 great enough genetic distance. These three species pairs are all allopatrically distributed. Due to  
442 their high levels of reproductive isolation (between 0.94-1.0), we expect that they would remain  
443 distinct upon secondary contact.

444         Of the 14 species pairs failing to meet the minimum total reproductive isolation  
445 threshold, 10 were allopatrically distributed. Because laboratory tests on reproductive isolation  
446 do not take geographic or environmental reproductive isolation into account, it is possible that  
447 allopatric pairs have not evolved a high degree of pre- or postzygotic isolation simply because  
448 they are genetically isolated by virtue of their distribution (i.e., geographic isolation is a strong  
449 barrier to gene flow, in itself). For example, the allopatric species pair of *D. americana* and *D.*  
450 *virilis* had relatively low genetic distance ( $D = 0.54$ ) and total isolation ( $T = 0.644$ ), but were  
451 never lumped under the MSC model ( $PPIL = 1.0$ ). Furthermore, detailed crossing experiments  
452 and QTL mapping have shown high levels of postmating, prezygotic isolation between these two  
453 species (Sweigart 2010), suggesting that genetic distance and total reproductive isolation may  
454 take more time to evolve relative to isolation identified under a coalescent framework. The other  
455 four species pairs with high PPIL but low reproductive isolation were all sympatric. These may  
456 represent cases where BPP is identifying incipient species with (potentially) low levels of  
457 ongoing or recent gene flow (Leaché et al. 2019), or cases where the amount of reproductive  
458 isolation may have been underestimated. For example, the sympatric species pair *D. lummei* – *D.*

459 *virilis* would be considered independent lineages under the MSC (PPIL = 1.0) despite having the  
460 lowest total isolation value in the entire dataset ( $T = 0.263$ ). However, at least one source  
461 (Heikkinen and Lumme 1991) reported high postzygotic isolation between the pair that was not  
462 reflected in either Coyne & Orr (1989, 1997) or Yukilevich (2012), and therefore was not  
463 included our calculation of total isolation for this species pair.

464         The remaining six species pairs in disagreement would be considered species based on  
465 reproductive isolation and genetic distance, but were not identified as independent lineages under  
466 the MSC. At face value, this would indicate these are reproductively isolated species pairs that  
467 cannot be identified as independently coalescing lineages. Of these, five are sympatric, and only  
468 one is allopatric. For the only allopatric pair (*D. mesophragmatic* – *D. pavani*), the PPIL value  
469 was barely below the 0.95 posterior probability cutoff (PPIL = 0.942), and probably represents a  
470 valid split under both methods. We recovered a range of PPIL values for the five sympatric  
471 species pairs (PPIL between 0.794-0.930). Additionally, we observed that for all the sympatric  
472 species pairs with high reproductive isolation but  $PPIL \leq 0.95$  one or both species in that pair  
473 have documented chromosomal inversions (Jha and Rahman 1973; Johnson 1985; Noor et al.  
474 2001).

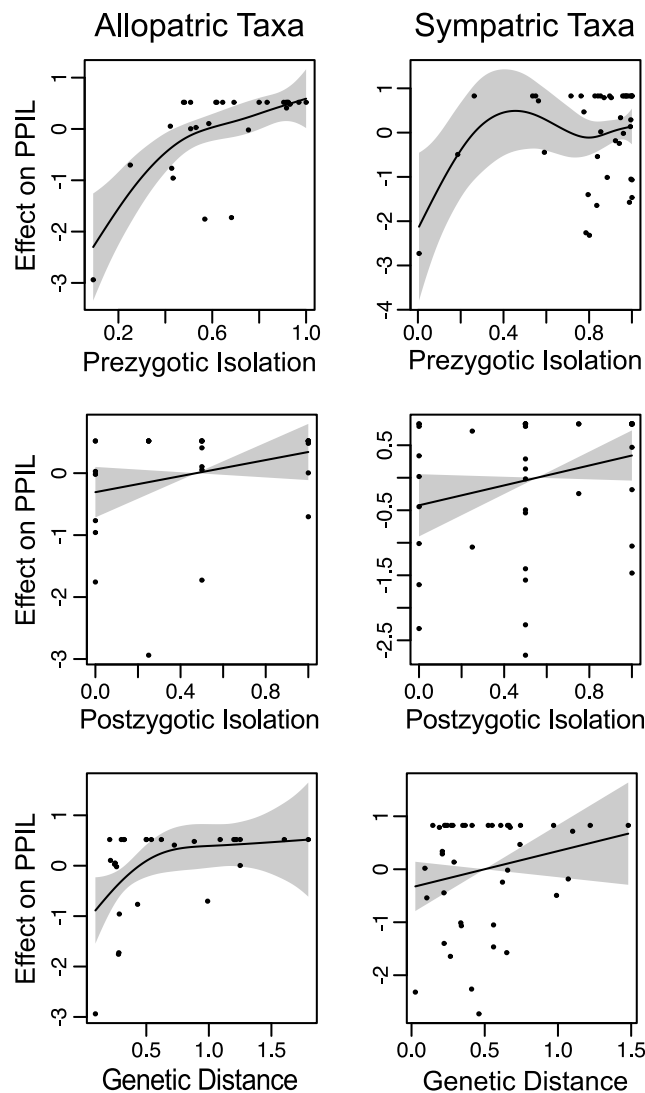
475         In principle, any of these discrepancies could be attributed to error in measures of  
476 reproductive isolation, genetic distance, or PPIL (e.g., poor fit of the MSC to the data).  
477 Widespread error in reproductive isolation measurements seems unlikely due to the controlled  
478 laboratory conditions under which these data were collected. However, extrinsic isolating  
479 mechanisms (e.g., ecological selection against hybrids) are not considered in a laboratory  
480 environment, and may represent a greater source of discordance between laboratory and nature  
481 than anything else. The amount of genetic distance could also have been overestimated or  
482 underestimated, but this seems unlikely to be a systematic issue given that we found a strong

483 correlation between Nei's electrophoretic distance and Jukes-Cantor genetic distance based on  
484 COII (Fig. S1). While it could be true that species pairs are not as distinct as the amount of  
485 genetic distance and reproductive isolation between them might suggest, it is also possible that  
486 we failed to delimit them due to a lack of statistical power, driven by a paucity of loci used to fit  
487 the MSC model. The lowest PPIL among reproductively isolated species (*D. algonquin* – *D.*  
488 *athabasca*; PPIL = 0.79) was based on sequence data from only one individual per species for  
489 two genes, which is among the lowest amount of data for any species pair in the entire dataset  
490 (Table S3). Discordance between methods in these cases may simply be driven by the amount of  
491 data available for analysis, rather than a true disagreement between methods. However,  
492 performance in BPP is good even with a small number of loci (Rannala and Yang 2017), and we  
493 found no relationship between the number of loci compared between species pairs and the PPIL  
494 value for that species pair (linear regression:  $R^2 = 0.002$ ; p-value = 0.61; Fig. S7).

495

#### 496 **Reproductive Isolation and GAMs**

497 We observed significant model improvement when specifying separate smoothers for sympatric  
498 and allopatric taxa, rather than considering all taxa jointly, for most explanatory variables (Table  
499 1). Based on these results, we subsequently explored the predictive power of three variables, and  
500 two combinations of reproductive isolation and genetic distance (see Materials & Methods), to  
501 predict PPIL values for sympatric and allopatric taxa separately. In allopatric taxa, we found that  
502 the combination model of prezygotic isolation + genetic distance best predicted PPIL value, with



503  
504  
505  
506  
507  
508

Figure 3) Fitted GAMs for allopatric (left) and sympatric (right) taxa. Note that the smooths are centered on zero to ensure model identifiability, meaning the y-axis is scaled relative to PPIL values but is not on the same scale (i.e., not from 0 – 1.0).

509 prezygotic isolation alone being found to be similarly plausible ( $\Delta AIC = 1.11$ ). This indicates  
510 that knowledge about prezygotic isolation, particularly in conjunction with genetic distance, is a  
511 strong predictor of PPIL values for allopatric species pairs. Although this makes sense at face  
512 value, it is interesting that prezygotic isolation represents a good predictor of speciation for taxa  
513 whose ranges do not overlap and therefore do not have the potential to mate in nature (i.e.,  
514 prezygotic isolation does not matter if taxa are allopatrically distributed). Furthermore, the fact

515 that the model of postzygotic isolation was the worst predictor of PPIL for allopatric species  
516 pairs ( $\Delta AIC = 14.8$ ) suggests the fate of hybrids between allopatric taxa, either through lab trials  
517 or secondary contact, says little about the independence of those lineages under the MSC.  
518 Models of genetic distance and postzygotic isolation + genetic distance were also poor predictors  
519 of PPIL in allopatric taxa. Taken together, this seems to imply that neither the time since two  
520 species became geographically isolated (inferred from genetic distance) nor levels of hybrid  
521 sterility/inviability are particularly indicative of lineage independence in allopatric taxa.

522 In sympatric taxa, we found that the model of prezygotic isolation alone was a better fit  
523 than all other explanatory variables, although the combination model of prezygotic isolation +  
524 genetic distance represented a similarly plausible alternative model ( $\Delta AIC = 0.206$ ). This agrees  
525 with the expectation that prezygotic isolation should be important in maintaining species  
526 boundaries in sympatry. When looking at sympatric taxa at low genetic distance ( $D < 0.5$ ), we  
527 found the combined model of prezygotic isolation + genetic distance model best predicted PPIL.  
528 This suggests that time since divergence (inferred from genetic distance), and not just the amount  
529 of prezygotic isolation, is an important indicator of lineage independence among closely related  
530 and recently diverged sympatric taxa. The model for postzygotic isolation was the worst  
531 predictor of PPIL for closely related sympatric taxa ( $\Delta AIC = 4.56$ ). Furthermore, in contrast to  
532 the finding from Coyne and Orr (1989, 1997) we find that prezygotic isolation is the best  
533 predictor of PPIL in closely related ( $D \leq 0.50$ ) allopatric *and* sympatric species pairs. Given that  
534 reinforcement, by definition, does not occur in allopatry (i.e., there are no unfit hybrids), this  
535 reaffirms the relative importance of prezygotic isolation throughout the speciation process, even  
536 in the absence of any postzygotic isolation.

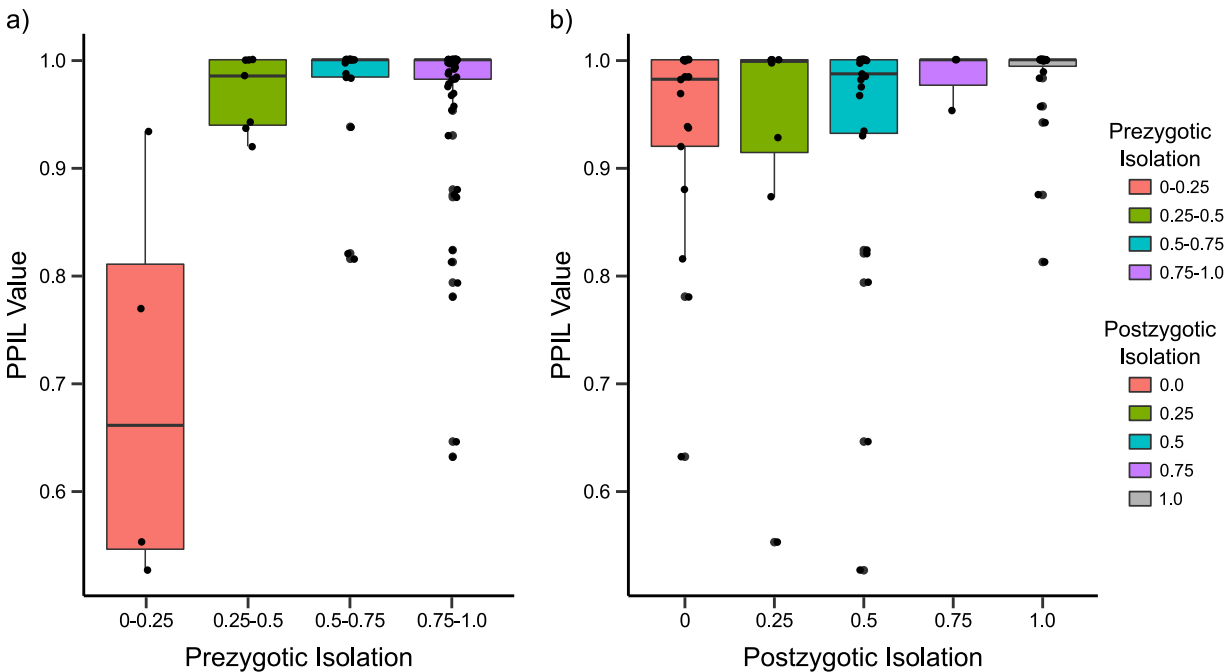
537 ANOVA tests revealed a significant difference between the level of prezygotic isolation  
538 and PPIL, but not between level of postzygotic isolation and PPIL, which recapitulates the



539 results of the GAM analysis. Furthermore, this shows that once a low level of prezygotic  
540 isolation has been obtained ( $\geq 0.25$ ) there is no statistical difference in PPIL values at higher  
541 levels of prezygotic isolation. Lastly, here was no statistically significant difference in PPIL for  
542 partially reproductively isolated species pairs (postzygotic isolation between 0.25-0.75) that do  
543 or do not show signature of Haldane's Rule. Although there is abundant support for the  
544 occurrence of Haldane's Rule in nature (Delph and Demuth 2016), sterility/infertility of the  
545 heterogametic sex does not seem to serve as a strong predictor of whether two lineages are  
546 identified as independent coalescent lineages.

#### 547 548 **Effect of the Prior**

549 Overall, we observed relatively consistent PPIL values across all prior settings (Table S4),  
550 indicating results from BPP analyses are not especially sensitive to the prior for either  $\theta$  or  $\tau$ . As  
551 we would expect, we observed larger differences between the prior and posterior distribution of  
552  $\theta$  and  $\tau$  for the uninformed priors than the informed priors (Fig. 4a). The informed priors  
553 allowed the MCMC to more widely explore topological space, both in terms of the number of  
554 species in a given topology and the frequency of proposing a unique topology. Uniformed priors  
555 are likely to depart strongly from empirically reasonable values for any particular dataset, which  
556 can constrain how well the MCMC explores areas of parameters space that may have high  
557 likelihoods (but very low prior probability). While we did not observe strong impacts on the  
558 mean number of species that were inferred for these data, the misspecified priors did lead to  
559 inflated estimates of certainty (Fig. 4b). Given the existing concerns regarding whether the MSC  
560 tends to oversplit lineages, an exaggerated increase in precision, and presumably accuracy in  
561 certain situations, would almost certainly exacerbate any innate complications of MSC species  
562 delimitation (see also Leaché et al. 2019 for an alternative approach to better characterizing  
563 uncertainty in MSC species delimitation).



564

565 Figure 4) Boxplots depicting the results of ANOVA test across different levels of pre- or postzygotic isolation.  
566 There was a significant difference for prezygotic isolation ( $p$ -value =  $2.01 \times 10^{-8}$ ), but not for postzygotic isolation  
567 ( $p$ -value = 0.343).

568

## 569 Implications for the practice of species delimitation

570 Here we present a detailed look at the difference between two species delimitation approaches in

571 a well-studied model organism, but our findings have implications beyond *Drosophila*

572 speciation. First, the overall amount of discordance between the two delimitation methods we

573 used here was much higher than we naïvely expected. If nominal taxa represent valid species,

574 then any model of species delimitation should recover them. Moreover, while the frequency of

575 speciation-collapse in nature is not yet well known, recent work has led to an increased

576 recognition that the evolutionary history of many species is reticulate in nature (i.e., periods of

577 gene flow during and/or after speciation; Burbrink and Gehara 2018; Marques et al. 2019) or

578 otherwise non-bifurcating (e.g., budding or anagenic speciation; Silvestro et al. 2018). Taken

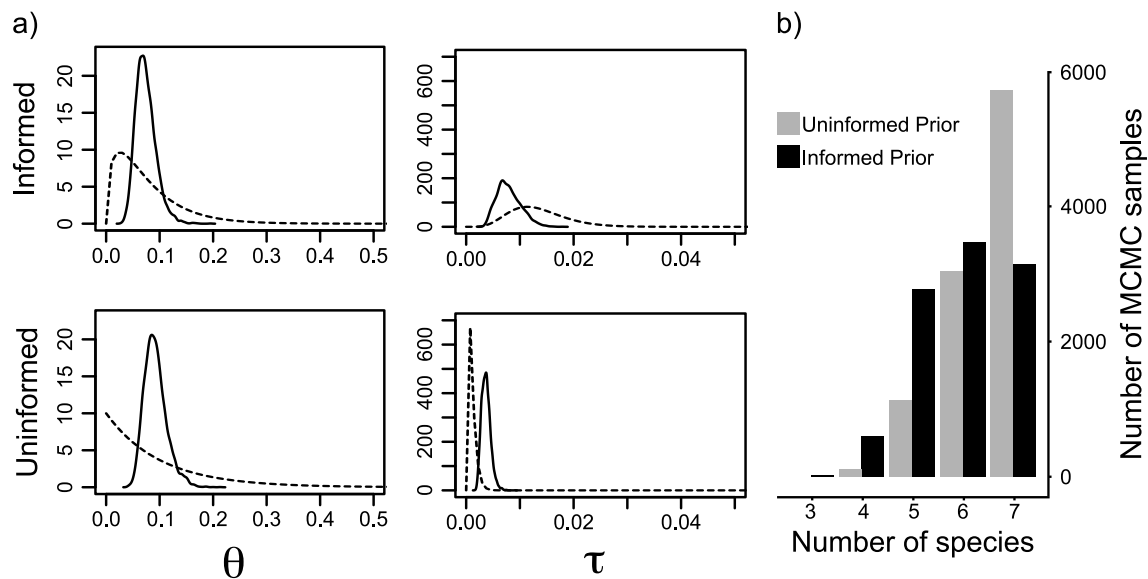
579 together, our results confirm that (if we take MSC delimitation as truth) reproductive isolation

580 does not need to be complete in order for lineages to be identifiable as independent.

581           Alternatively, the MSC model may not be able to pick up on reproductive isolating  
582 mechanisms that are specific to particular regions of the genome (i.e., genomic islands of  
583 divergence; Wolf and Ellegren 2016). In this study, we found that for sympatric species with  
584 high reproductive isolation, but low PPIL value, chromosomal rearrangement could be  
585 implicated in the maintenance of species boundaries. Moreover, while chromosomal  
586 rearrangements have long been associated with reproductive isolation and speciation (Rieseberg  
587 2001; Campbell et al. 2018), to our knowledge have not been formally investigated under a MSC  
588 model. Supposing chromosomal inversions drive a rapid increase in reproductive isolation, the  
589 effects of localized genomic islands of divergence may not extend throughout the genome or be  
590 reflected in the few loci being used for delimitation. Although selecting a number of genes from  
591 a heterogeneous gene pool will often provide a useful approximation of species limits, if the only  
592 difference between species are relatively small genomic islands of divergence (e.g.,  
593 chromosomal inversions) the results from MSC delimitation may be inconsistent with the levels  
594 of reproductive isolated observed in nature.

595           This study also allowed us to reveal novel insight into the speciation process. Most  
596 notably, we found that levels of prezygotic isolation were consistently related to PPIL, and found  
597 no indication that postzygotic isolation was similarly informative. Although quantification of  
598 pre- and postzygotic isolation are not readily available outside of model organisms, the  
599 implications of this finding potentially extends beyond the lab. Specifically, they may indicate  
600 that prezygotic isolation evolves earlier in the speciation process and rapidly leads to identifiable  
601 independent lineages, even under a relatively small amount of assortative mating (prezygotic  
602 isolation  $\geq 0.25$ ). This may provide some insight into why hybridization appears to be more  
603 common in nature than was previously appreciated (Taylor and Larson 2019). As long as most of  
604 the gene flow is within species, the boundaries are maintained and the speciation process in not

605 disrupted. However, further work is needed to move beyond such speculation. Overall, our  
 606 results highlight the value of considering the processes that underlie speciation, in conjunction  
 607 with the assumptions of each species delimitation method employed, as a fruitful means for  
 608 understanding conflicts (see also Barley et al. 2018; Smith and Carstens 2018).



609 Figure 5a) This figure compares relationships between prior (dashed) and posterior (solid) gamma distributions of  $\theta$   
 610 and  $\tau$  using informed (top) and uninformed (bottom) prior distributions for the Montium group. Plots for all other  
 611 species groups can be found in the Supplementary Material (Fig. S8). b) Shows the difference between the number  
 612 of species delimited under uninformed (gray) and informed (black) prior setting out of the total 10,000 MCMC  
 613 samples per run.  
 614  
 615

## 616 CONCLUSION

617 While reproductive isolation has long been a cornerstone of speciation research, it can rarely be  
 618 used as an operational delimitation method because collecting such data is not tractable for most  
 619 systems. Likewise, many practitioners of model-based species delimitation do not take  
 620 reproductive isolation into account because the data are unavailable. In this study, we focus on a  
 621 *Drosophila* dataset that is uniquely suited to address this divide in how species are understood.  
 622 Here, we formally investigated how species boundaries based on reproductive isolation compare  
 623 to model based species delimitation. We found that model based and reproductive isolation based  
 624 methods agreed on ~77% of species delimitations. MSC consistently delimited more species

625 pairs (~17%), while 6% of the species pairs were delimited based on reproductive isolation and  
626 genetic distance alone. However, because our study design does not allow for within species  
627 population structure to be misinterpreted as species boundaries, our results may provide a  
628 somewhat optimistic view of MSC performance. Additionally, by using a set of informed and  
629 uninformed prior settings we were able to qualitatively assess the accuracy and precision of MSC  
630 species delimitation, as implemented by BPP. In accordance with earlier studies, BPP appears to  
631 be relatively robust to prior misspecification, although we do find evidence that prior  
632 misspecification can lead to overconfidence in estimates of species boundaries. Lastly, based on  
633 predictive GAM models, we found that the amount of postzygotic isolation between species pairs  
634 was relatively uninformative as to whether or not those lineages can be recognized as  
635 independent under the coalescent. We are not questioning that postzygotic isolation acts as a  
636 strong reproductive barrier, certainly it does. However, it seems that prezygotic isolation  
637 evolving early in the speciation process is closely tied to recognizably independent lineages. As  
638 such, prezygotic isolation may be relatively more important in the incipient stages of speciation,  
639 and possibly to the process as a whole.

640

#### 641 **SUPPLEMENTARY MATERIAL**

642 Data available from the Dryad Digital Repository: XXXXXX

643

#### 644 **FUNDING**

645 This work was supported by the US National Science Foundation (grant numbers DEB 1354506,  
646 DEB 1754350, and DBI 1356796). This work was also supported by computational resources  
647 provided by the University of Hawai'i Cyberinfrastructure group, as well as funding from the  
648 Arnold and Mabel Beckman Foundation.

649 **REFERENCES**

- 650 Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on*  
651 *Automatic Control*. 19:716–723.
- 652 Barley A.J., Brown J.M., Thomson R.C. 2018. Impact of Model Violations on the Inference of  
653 Species Boundaries Under the Multispecies Coalescent. *Systematic Biology*. 67:269–284.
- 654 Burbrink F.T., Gehara M. 2018. The Biogeography of Deep Time Phylogenetic Reticulation.  
655 *Systematic Biology*. 67:743–755.
- 656 Burnham K.P., Anderson D.R., Burnham K.P. 2002. Model selection and multimodel inference:  
657 a practical information-theoretic approach. New York: Springer.
- 658 Camargo A., Morando M., Avila L.J., Sites J.W. 2012. Species delimitation with ABC and other  
659 coalescent-based methods: a test of accuracy with simulations and an empirical example  
660 with lizards of the *Liolaemus darwini* complex (Squamata: Liolaemidae): species  
661 delimitation with ABC. *Evolution*. 66:2834–2849.
- 662 Campbell C.R., Poelstra J.W., Yoder A.D. 2018. What is Speciation Genomics? The roles of  
663 ecology, gene flow, and genomic architecture in the formation of species. *Biological*  
664 *Journal of the Linnean Society*. 124:561–583.
- 665 Castillo D.M. 2017. Factors contributing to the accumulation of reproductive isolation: A mixed  
666 model approach. *Ecology and Evolution*. 7:5808–5820.
- 667 Chambers E.A., Hillis D.M. 2019. The Multispecies Coalescent Over-Splits Species in the Case  
668 of Geographically Widespread Taxa. *Systematic Biology*.
- 669 Charlesworth B. 2009. Fundamental concepts in genetics: Effective population size and patterns  
670 of molecular evolution and variation. *Nature Reviews Genetics*. 10:195–205.
- 671 Coyne J.A., Orr H.A. 1989. Patterns of Speciation in *Drosophila*. *Evolution*. 43:362.
- 672 Coyne J.A., Orr H.A. 1997. “Patterns of Speciation in *Drosophila*” Revisited. *Evolution*. 51:295.
- 673 Coyne J.A., Orr H.A. 2004. Speciation. Sunderland, Mass: Sinauer Associates.
- 674 Cutter A.D. 2008. Divergence Times in *Caenorhabditis* and *Drosophila* Inferred from Direct  
675 Estimates of the Neutral Mutation Rate. *Molecular Biology and Evolution*. 25:778–786.
- 676 Delph L.F., Demuth J.P. 2016. Haldane’s Rule: Genetic Bases and Their Empirical Support.  
677 *Journal of Heredity*. 107:383–391.
- 678 Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high  
679 throughput. *Nucleic Acids Research*. 32:1792–1797.
- 680 Eyre-Walker A., Keightley P.D., Smith N.G.C., Gaffney D. 2002. Quantifying the slightly  
681 deleterious mutation model of molecular evolution. *Mol. Biol. Evol.* 19:2142–2149.
- 682 Felsenstein J. 2004. Inferring phylogenies. Sunderland, Mass: Sinauer Associates.
- 683 Fox J., Weisberg S. 2011. An R Companion to Applied Regression. Thousand Oaks CA: Sage.
- 684 Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based  
685 species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*. 27:480–  
686 488.
- 687 Haag-Liautard C., Dorris M., Maside X., Macaskill S., Halligan D.L., Charlesworth B.,  
688 Keightley P.D. 2007. Direct estimation of per nucleotide and genomic deleterious  
689 mutation rates in *Drosophila*. *Nature*. 445:82–85.
- 690 Haldane J.B.S. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*.  
691 12:101–109.
- 692 Heikkinen E., Lumme J. 1991. Sterility of male and female hybrids of *Drosophila virilis* and  
693 *Drosophila lummei*. *Heredity*. 67:1.
- 694 Hey J., Nielsen R. 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates  
695 and Divergence Time, With Applications to the Divergence of *Drosophila*  
696 *pseudoobscura* and *D. persimilis*. *Genetics*. 167:747–760.

- 697 Jha A.P., Rahman S.M. 1973. On crossing between *Drosophila bipectinata* and *Drosophila*  
698 *malerkotliana*. Cytologia (Tokyo). 38:425–436.
- 699 Johnson D.L.E. 1985. Genetic differentiation in the *Drosophila athabasca* complex. Evolution.  
700 39:467–472.
- 701 Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A.,  
702 Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012.  
703 Geneious Basic: An integrated and extendable desktop software platform for the  
704 organization and analysis of sequence data. Bioinformatics. 28:1647–1649.
- 705 Keightley P.D., Ness R.W., Halligan D.L., Haddrill P.R. 2014. Estimation of the Spontaneous  
706 Mutation Rate per Nucleotide Site in a *Drosophila melanogaster* Full-Sib Family.  
707 Genetics. 196:313–320.
- 708 Keightley P.D., Trivedi U., Thomson M., Oliver F., Kumar S., Blaxter M.L. 2009. Analysis of  
709 the genome sequences of three *Drosophila melanogaster* spontaneous mutation  
710 accumulation lines. Genome Research. 19:1195–1201.
- 711 Leaché A.D., Fujita M.K. 2010. Bayesian species delimitation in West African forest geckos  
712 (*Hemidactylus fasciatus*). Proceedings of the Royal Society B: Biological Sciences.  
713 277:3071–3077.
- 714 Leaché A.D., Zhu T., Rannala B., Yang Z. 2019. The Spectre of Too Many Species. Systematic  
715 Biology. 68:168–181.
- 716 Legrand D., Tenaillon M.I., Matyot P., Gerlach J., Lachaise D., Cariou M.-L. 2009. Species-  
717 Wide Genetic Variation and Demographic History of *Drosophila sechellia*, a Species  
718 Lacking Population Structure. Genetics. 182:1197–1206.
- 719 van der Linde K., Houle D. 2008. A supertree analysis and literature review of the genus  
720 *Drosophila* and closely related genera (Diptera, Drosophilidae). Insect Systematics &  
721 Evolution. 39:241–267.
- 722 Luo A., Ling C., Ho S.Y., Zhu C. 2018. Comparison of Methods for Molecular Species  
723 Delimitation across a Range of Speciation Scenarios. Systematic Biology.
- 724 Marques D.A., Meier J.I., Seehausen O. 2019. A Combinatorial View on Speciation and  
725 Adaptive Radiation. Trends in Ecology & Evolution. 34:531–544.
- 726 Mayr E. 1963. Animal species and evolution. Belknap Press of Harvard University Press.
- 727 Noor M.A.F., Grams K.L., Bertucci L.A., Reiland J. 2001. Chromosomal inversions and the  
728 reproductive isolation of species. Proceedings of the National Academy of Sciences.  
729 98:12084–12088.
- 730 Obbard D.J., Maclennan J., Kim K.-W., Rambaut A., O’Grady P.M., Jiggins F.M. 2012.  
731 Estimating Divergence Dates and Substitution Rates in the *Drosophila* Phylogeny.  
732 Molecular Biology and Evolution. 29:3459–3473.
- 733 O’Grady P.M., DeSalle R. 2018. Phylogeny of the Genus *Drosophila*. Genetics. 209:1–25.
- 734 Pascual M., Chapuis M.P., Mestres F., Balanyà J., Huey R.B., Gilchrist G.W., Serra L., Estoup  
735 A. 2007. Introduction history of *Drosophila subobscura* in the New World: a  
736 microsatellite-based survey using ABC methods. Molecular Ecology. 16:3069–3083.
- 737 de Queiroz K. 1998. The General Lineage Concept of Species, Species Criteria, and the  
738 Process of Speciation. In: Howard D.J., Berlocher S.H., editors. Endless Forms: Species  
739 and Speciation. Oxford University Press. p. 57–75.
- 740 de Queiroz K. 2007. Species Concepts and Species Delimitation. Systematic Biology. 56:879–  
741 886.
- 742 R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna,  
743 Austria: R Foundation for Statistical Computing.



- 744 Rannala B., Yang Z. 2003. Bayes estimation of species divergence times and ancestral  
745 population sizes using DNA sequences from multiple loci. *Genetics*. 164:1645–1656.
- 746 Rannala B., Yang Z. 2017. Efficient Bayesian species tree inference under the multispecies  
747 coalescent. *Systematic biology*. 66:823–842.
- 748 Reid N.M., Hird S.M., Brown J.M., Pelletier T.A., McVay J.D., Satler J.D., Carstens B.C. 2014.  
749 Poor Fit to the Multispecies Coalescent is Widely Detectable in Empirical Data.  
750 *Systematic Biology*. 63:322–333.
- 751 Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other  
752 things): phytools: R package. *Methods in Ecology and Evolution*. 3:217–223.
- 753 Rieseberg L.H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology &*  
754 *Evolution*. 16:351–358.
- 755 Schrider D.R., Houle D., Lynch M., Hahn M.W. 2013. Rates and Genomic Consequences of  
756 Spontaneous Mutational Events in *Drosophila melanogaster*. *Genetics*. 194:937–954.
- 757 Silvestro D., Warnock R.C.M., Gavryushkina A., Stadler T. 2018. Closing the gap between  
758 palaeontological and neontological speciation and extinction rate estimates. *Nature*  
759 *Communications*. 9.
- 760 Simpson G.G. 1951. The Species Concept. *Evolution*. 5:285–298.
- 761 Smith G., Lohse K., Etges W.J., Ritchie M.G. 2012. Model-based comparisons of  
762 phylogeographic scenarios resolve the intraspecific divergence of cactophilic *Drosophila*  
763 *mojavensis*: Phylogeography of *Drosophila mojavensis*. *Molecular Ecology*. 21:3293–  
764 3307.
- 765 Smith M.L., Carstens B.C. 2018. Disentangling the process of speciation using machine learning.  
766 bioRxiv.
- 767 Sukumaran J., Holder M.T. 2010. DendroPy: a Python library for phylogenetic computing.  
768 *Bioinformatics*. 26:1569–1571.
- 769 Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species.  
770 *Proceedings of the National Academy of Sciences*. 114:1607–1612.
- 771 Sweigart A.L. 2010. The Genetics of Postmating, Prezygotic Reproductive Isolation Between  
772 *Drosophila virilis* and *D. americana*. *Genetics*. 184:401–410.
- 773 Tamura K. 2003. Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation  
774 Clocks. *Molecular Biology and Evolution*. 21:36–44.
- 775 Taylor S.A., Larson E.L. 2019. Insights from genomes into the evolutionary importance and  
776 prevalence of hybridization in nature. *Nature Ecology & Evolution*. 3:170–177.
- 777 Wall J.D., Andolfatto P., Przeworski M. 2002. Testing models of selection and demography in  
778 *Drosophila simulans*. *Genetics*. 162:203–216.
- 779 Watada M., Matsumoto M., Kondo M., Kimura M.T. 2011. Taxonomic study of the *Drosophila*  
780 *auraria* species complex (Diptera: Drosophilidae) with description of a new species:  
781 *Drosophila auraria* species complex. *Entomological Science*. 14:392–398.
- 782 Wiley E.O. 1978. The Evolutionary Species Concept Reconsidered. *Systematic Zoology*. 27:17.
- 783 Willis S.C. 2017. One species or four? Yes!...and, no. Or, arbitrary assignment of lineages to  
784 species obscures the diversification processes of Neotropical fishes. *PLOS ONE*.  
785 12:e0172349.
- 786 Wolf J.B.W., Ellegren H. 2016. Making sense of genomic islands of differentiation in light of  
787 speciation. *Nature Reviews Genetics*. 18:87–100.
- 788 Wood S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation  
789 of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*.  
790 73:3–36.



- 791 Yang Y., Hou Z.-C., Qian Y.-H., Kang H., Zeng Q.-T. 2012. Increasing the data size to  
792 accurately reconstruct the phylogenetic relationships between nine subgroups of the  
793 *Drosophila melanogaster* species group (Drosophilidae, Diptera). *Molecular*  
794 *Phylogenetics and Evolution*. 62:214–223.
- 795 Yang Z. 2015. The BPP program for species tree estimation and species delimitation. *Current*  
796 *Zoology*. 61:854–865.
- 797 Yang Z., Rannala B. 2014. Unguided Species Delimitation Using DNA Sequence Data from  
798 Multiple Loci. *Molecular Biology and Evolution*. 31:3125–3135.
- 799 Yi S., Bachtrog D., Charlesworth B. 2003. A survey of chromosomal and nucleotide sequence  
800 variation in *Drosophila miranda*. *Genetics*. 164:1369–1381.
- 801 Yukilevich R. 2012. Asymmetrical patterns of speciation uniquely support reinforcement in  
802 *Drosophila*: asymmetrical patterns of reinforcement speciation. *Evolution*. 66:1430–  
803 1446.