

1 **A framework for estimating the effects of sequential reproductive barriers:**
2 **implementation using Bayesian models with field data from cryptic species**

3 **Short title:** estimation of sequential reproductive barriers

4 Jean Peccoud^{1,2,*}, David R. J. Pleydell^{1,3,*}, Nicolas Sauvion¹

5 ¹BGPI, Univ Montpellier, INRA, CIRAD, Montpellier SupAgro, Montpellier, France

6 ²Current address: UMR Écologie et Biologie des Interactions, CNRS, Université de
7 Poitiers, Poitiers, France.

8 ³Current address: UMR Animal, Santé, Territoires, Risques et Écosystèmes, INRA,
9 CIRAD, Montpellier SupAgro, Université de Montpellier, Montpellier, France.

10 * These authors contributed equally to this work.

11

12 **Corresponding author:** Nicolas Sauvion, Tel: +33 4 99 62 48 21 Email:
13 nicolas.sauvion@inra.fr

14 **Acknowledgements:** We thank René Rieu for his advice on spermatophore
15 extraction, and to Gaël Thébaud, Gérard Labonne and François Bonnot for helpful
16 comments on the study and drafts of this article. We thank Josiane Peyre and Patrick
17 Limon for their help in genotyping. This work utilized computing resources from
18 INRA's MIGALE cluster (<http://migale.jouy.inra.fr>) and benefitted from the
19 assistance of Eric Montaudon and Véronique Martin. Part of this work was funded
20 by INRA grant SDIPS (Speciation and molecular Diagnosis of Insect Pest Species
21 complexes).

22 **Author contributions:** JP and NS initiated the study and obtained biological data. JP
23 and DRJP developed the porosity-based approach. DRJP conceived the Bayesian
24 implementation and code. JP, DRJP and NS wrote the manuscript.

25 **Data availability:** Mitochondrial sequence data will be available at Genbank, source
26 code is available at xxx.

27 **Abstract**

28 Determining how reproductive barriers modulate gene flow between populations
29 represents a major step towards understanding the factors shaping the course of
30 speciation. Although many indices quantifying reproductive isolation (RI) have been
31 proposed, they do not permit the quantification of cross direction-specific RI under
32 varying species frequencies and over arbitrary sequences of barriers. Furthermore,
33 techniques quantifying associated uncertainties are lacking, and statistical methods
34 unrelated to biological process are still preferred for obtaining confidence intervals
35 and p-values. To address these shortcomings, we provide new RI indices that model
36 changes in gene flow for both directions of hybridization, and we implement them in
37 a Bayesian model. We use this model to quantify RI between two species of the
38 psyllid *Cacopsylla pruni* based on field genotypic data for mating individuals,
39 inseminated spermatophores and progeny. The results showed that pre-
40 insemination isolation was strong, mildly asymmetric and undistinguishably
41 different between study sites despite large differences in species frequencies; that
42 post-insemination isolation strongly affected the more common hybrid type; and
43 that cumulative isolation was close to complete. In the light of these results, we
44 discuss how these developments can strengthen comparative RI studies.

45 **Keywords:** reproductive isolation, hybridization, Bayesian modeling, finite mixture
46 models, asymmetrical isolation, gene flow.

47

48 **Introduction**

49 Speciation involves the build-up of reproductive isolation (RI) at several key parts of
50 the populations' life cycles, which are referred to as reproductive barriers.
51 Understanding how these barriers act in conjunction to reduce gene flow and permit
52 the divergence of populations into species has been an important goal of speciation
53 research (Coyne and Orr 2004; Sobel et al. 2010; Butlin et al. 2012). As a result, the
54 last fifteen years have seen a burgeoning of methods to estimate the strength of
55 reproductive barriers from field data on natural populations (Ramsey et al. 2003;
56 Malausa et al. 2005; Martin and Willis 2007; Matsubayashi and Katakura 2009;
57 Sanchez-Guillen et al. 2012; Sobel and Streisfeld 2015; Pombi et al. 2017). Field
58 estimates indeed provide the most pertinent results to help identify local factors
59 affecting the course of speciation (Nosil et al. 2009; Via 2009; Sobel et al. 2010;
60 Butlin et al. 2012).

61 This objective requires that RI estimates represent evolutionary relevant quantities,
62 mainly potential gene flow, whilst correcting for differences in species frequencies
63 that do not reflect phenotypic variations – effects we collectively refer to as
64 “contingency”. In the progression toward this goal, many indices to quantify RI have
65 been developed [reviewed in Sobel and Chen (2014)]. Key developments include

66 formulas to quantify cumulative RI over sequential reproductive barriers (Ramsey
67 et al. 2003); corrections for unequal species frequencies and allochrony (Martin and
68 Willis 2007); and the integration of these developments into RI indices that
69 maintain a linear relation to the probability of gene flow – a desirable property
70 when comparing populations and species (Sobel and Chen 2014).

71 Despite this growing diversity and sophistication of RI indices, and of the studies
72 using them, two deficiencies of current methods remain apparent. First, although RI
73 is commonly asymmetrical (e.g., Lowry et al. 2008; Matsubayashi and Katakura
74 2009; Sanchez-Guillen et al. 2012; Brys et al. 2014; Kaufmann et al. 2017; Martin et
75 al. 2017), we lack indices that can estimate directional (cross-type specific) RI over
76 an arbitrary combination of reproductive barriers, while controlling for
77 contingency. Second, speciation research would benefit from more studies reporting
78 and discussing uncertainty in RI (e.g., Merrill et al. 2011; Lackey and Boughman
79 2017). The RI literature is dominated by the discussion of point estimates for which
80 there exists a lack of associated uncertainty measures. Thus, it remains difficult to
81 demonstrate whether apparent differences in RI (observed between different
82 barriers, sub-populations or species) reflect real phenotypic differences or merely
83 sampling error. Accordingly, practitioners seeking a richer statistical analysis -
84 involving confidence intervals or significance tests for example - have been
85 constrained to adopt less biologically-motivated indices such as those provided by
86 generalized linear models (Takami et al. 2007; Polacik and Reichard 2011; Falk et al.
87 2012; Peccoud et al. 2014; Kostyun and Moyle 2017).

88 Why a more complete statistical framework for RI estimation has not emerged may
89 partly stem from the fact that the calculation of RI is frequently complexified by the
90 need to correct for contingency and the effects of reproductive barriers not under
91 scrutiny, or to combine the effects of several barriers (Sobel and Chen 2014).
92 Accounting for asymmetry in RI would further complexify existing formulas and
93 pose a substantial challenge regarding the construction of confidence intervals and
94 significance tests for these indices.
95 We suggest that these issues can be resolved by focusing attention on estimating the
96 probabilities of gene flow – rather than RI *per se* – induced by both within- and
97 between-species crosses. Focusing on the probabilities of gene flow facilitates
98 statistical estimation, from field data, of contingency-independent RI indices (in
99 both cross directions) at any reproductive barrier or over any arbitrary sequence of
100 barriers. Moreover, this approach naturally lends itself to Bayesian uncertainty
101 analysis. In other branches of ecology and evolution, Bayesian techniques have long
102 been popular for numerous reasons, including: they provide a natural paradigm to
103 account for multiple sources of uncertainty; they facilitate the incorporation of prior
104 knowledge; they are applicable to a wide variety of models; and inference based on
105 posterior distributions of model parameters is easy and intuitive (Gelman et al.
106 1995; Clark 2005; Cressie et al. 2009; Beaumont 2010; Hoban et al. 2012; Gompert
107 et al. 2017). We illustrate these benefits with a Bayesian model designed to quantify
108 the weight of evidence for spatial heterogeneity in RI using genetic data from
109 natural populations of the psyllid *Cacopsylla pruni*.

110 Methods

111 *Modeling sequential reproductive barriers*

112 Consider two species A and B interacting at a reproductive barrier. Let G_{XY} denote
113 the probability that an individual sampled from the next generation comes from an
114 $X \times Y$ cross (with $X, Y \in \{A, B\}$ and the maternal species always noted first) in the
115 absence of further isolation after the barrier. Thus, G_{XY} is the potential gene flow
116 (which we may simply call “gene flow” afterwards) induced by $X \times Y$ crosses. Let
117 $\mathbf{G} = \{G_{AA}, G_{BB}, G_{AB}, G_{BA}\}$ be the set of all such proportions, which sum to one.

118 Estimating RI as the decrease of interspecific gene flow (Sobel and Chen 2014)
119 requires a measure of gene flow that is independent of contingency. We call these
120 contingency-independent gene flow rates “barrier porosities” to convey that they
121 solely depend on phenotypic differences expressed at the barrier. We denote barrier
122 porosities as $\boldsymbol{\beta} = \{\beta_{AA}, \beta_{BB}, \beta_{AB}, \beta_{BA}\}$, these sum to one and each β_{XY} element equals 1/4
123 in the absence of RI.

124 The ratio of porosity, over its null expectation when RI=0, indicates the strength of
125 RI at a barrier (Sobel and Chen 2014). Thus, a bidirectional RI index that considers
126 both hybrid cross-types is

$$127 \quad RI = 1 - 2(\beta_{AB} + \beta_{BA}) \quad (1)$$

128 and the RI affecting just one hybrid cross-type ($X \neq Y$) is:

129 $RI_{XY} = 1 - 4\beta_{XY}.$ (2)

130 These indices vary linearly with gene flow, take value zero when porosities are 1/4,
131 and take value one when porosity to hybridization is zero.

132 Directional RI indices allow between cross-type differences (asymmetry) in RI to be
133 quantified as

134 $\Delta = RI_{AB} - RI_{BA}.$

135 Given the simplicity of these developments, the main modeling task is to establish
136 the relationships between gene flow G and barrier porosities β . To this aim, we
137 introduce “null gene flow” $E_0[G_{XY}]$ to denote gene flow in the absence of RI at the
138 studied barrier. $E_0[G_{XY}]$ can be visualized as the flow of genes going through the
139 previous barrier and arriving at the focal barrier. Table 1 provides examples of
140 $E_0[G_{XY}]$ and G_{XY} for different sources of RI.

141 At the postzygotic level, the relative frequency of XY genotypes in the progeny, G_{XY} , is
142 proportional to their frequency before the barrier, $E_0[G_{XY}]$, multiplied by their
143 probability to survive (or pass) through the barrier (defined as S_{XY}):

144 $G_{XY} \propto E_0[G_{XY}]S_{XY}.$ (3)

145 Under equal species frequencies and random mating, $G_{XY} = \beta_{XY}$ by definition and
146 $E_0[G_{XY}] = 1/4$. Assuming that survival S_{XY} is constant, and is therefore independent of
147 species frequencies, the above equation translates to:

148 $\beta_{XY} \propto S_{XY}$

149 where 1/4 was dropped as a constant in the proportionality relationship.

150 Substituting S_{XY} in equation 3 yields:

151
$$G_{XY} \propto E_0[G_{XY}] \beta_{XY}. \quad (4)$$

152 Postzygotic barrier porosities are therefore proportional to the ratio of potential
153 over null gene flow – a metric that enables RI quantification when null expectations
154 (elements of $E_0[G]$) are unequal [see appendix D of Sobel and Chen (2014)].

155 For total gene flow to equal one, equation 4 requires normalization:

156
$$G_{XY} = \frac{E_0[G_{XY}] \beta_{XY}}{\sum_{X \in \{A,B\}} \sum_{Y \in \{A,B\}} E_0[G_{XY}] \beta_{XY}}. \quad (5)$$

157 Equation 5 satisfies that when all null gene flows equal 1/4 (equal species
158 frequencies), the porosity β_{XY} equals gene flow G_{XY} . Conversely, in the absence of RI
159 (β_{XY} is 1/4 for all combinations), gene flows equal null gene flows.

160 Given estimates of gene flow before and after a barrier, the porosities can be
161 recovered by rearranging and normalizing equation 4 so that element of β sum to 1,
162 hence:

163
$$\beta_{XY} = \frac{G_{XY}/E_0[G_{XY}]}{\sum_{X \in \{A,B\}} \sum_{Y \in \{A,B\}} (G_{XY}/E_0[G_{XY}])}. \quad (6)$$

164 Contrarily to postzygotic barriers that increase progeny mortality, prezygotic
165 barriers do not usually incur a fitness cost to parents. Hence, prezygotic isolation
166 must be modeled in such a way that it does not directly affect fitness. To do so, we
167 express the proportion of XY zygotes (that we expect if no isolation exists after the
168 studied barrier) among those having a mother from species X :

169
$$\frac{G_{XY}}{G_{XA} + G_{XB}} = \frac{E_0[G_{XY}]\beta_{XY}}{E_0[G_{XA}]\beta_{XA} + E_0[G_{XB}]\beta_{XB}}. \quad (7)$$

170 In order to derive G_{XY} , the relative contribution of species X females to the next
171 generation ($G_{XA} + G_{XB}$) must be specified. If, at the focal barrier, female reproductive
172 success does not vary between the species, then $G_{XA} + G_{XB}$ is the frequencies of
173 species X in females, which we call f_x , and equation 7 becomes:

174
$$G_{XY} = f_x \frac{E_0[G_{XY}]\beta_{XY}}{E_0[G_{XA}]\beta_{XA} + E_0[G_{XB}]\beta_{XB}}. \quad (8)$$

175 Barrier porosities can be obtained from gene flows by rearranging equation 7 into
176 (proof not shown):

177

178
$$\frac{\beta_{XY}}{\beta_{XA} + \beta_{XB}} = \frac{G_{XY}/E_0[G_{XY}]}{G_{XA}/E_0[G_{XA}] + G_{XB}/E_0[G_{XB}]}, \quad (9)$$

179 and specifying $\beta_{XA} + \beta_{XB}$ appropriately. If female reproductive success can be
180 assumed equal between species, then $\beta_{XA} + \beta_{XB} = 1/2$, so:

181
$$\beta_{XY} = \frac{1}{2} \frac{G_{XY}/E_0[G_{XY}]}{G_{XA}/E_0[G_{XA}] + G_{XB}/E_0[G_{XB}]}. \quad (10)$$

182 Equation 7, and its by-products, assume that the fraction of XY zygotes contributed
183 by females of species X is proportional to its null-expected value in the absence of RI,
184 $E_0[G_{XY}]$. This implies that the probability of hybridization per interspecific
185 encounter, represented by the ratio $G_{XY}/E_0[G_{XY}]$, does not depend on species
186 frequencies, and reflects the barrier porosity β_{XY} . If the assumption of equal
187 reproductive success between females is not warranted, alternative formulations for

188 equations 8 and 10 may be desirable. Such developments should be tailored to the
189 specifics of the biological system and are beyond the scope of the current work.

190 Once obtained, barrier porosities can be used to model a sequence of b barriers
191 with porosities $\beta^1 \dots \beta^b$. The product of these porosities is proportional to the
192 probability that genes from two parents flow through all these barriers to eventually
193 produce an offspring. The combined porosity of these barriers to $X \times Y$ gene flow is
194 thus given by:

$$195 \quad \beta_{XY}^{1:b} = \frac{\prod_{i=1}^b \beta_{XY}^i}{\sum_{X \in \{A,B\}} \sum_{Y \in \{A,B\}} \prod_{i=1}^b \beta_{XY}^i}, \quad (11)$$

196 whose denominator ensures that the combined porosities of all four XY
197 combinations sum to one.

198 Equations 5 and 8 permit barrier porosities β_{XY} , hence RI, to be estimated via
199 statistical techniques that confront modelled gene flows G_{XY} with data collected at
200 different points of the reproductive cycle. We will demonstrate this approach with a
201 Bayesian implementation. Alternatively, a simpler approach would use equations 6
202 or 10 to obtain point estimates of barrier porosities by specifying G_{XY} according to
203 observations (examples given in Table 1).

204 *Study model*

205 Our model system, *Cacopsylla pruni* Scopoli (Sternorrhyncha: Psyllidae), includes
206 two unnamed cryptic species which are strongly genetically divergent but have yet
207 to show ecological or morphological differences (Sauvion et al. 2007; Peccoud et al.
208 2013). These species co-occur at several sites in Southern France (Sauvion et al.

209 2007) on shrubs of genus *Prunus*, on which the insects feed, reproduce and die in
210 spring (Figure 1). Progeny reach adulthood after approximately 2 months, migrate
211 shortly after to conifers for overwintering and return to *Prunus* in early spring to
212 mate.

213 Rearing *C. pruni* in controlled conditions has proven extremely difficult (Jarausch
214 and Jarausch 2016). However, the non-overlapping generations and co-occurrence
215 of the *C. pruni* species at several sites make them good candidates for field-based
216 estimates of RI within their life-cycle (Figure 1). To this aim, we genotyped mating
217 adults, inseminated spermatophores and progeny as species *A* or *B* or as hybrids.

218 *Sample collection and species assignment*

219 Psyllids were collected in spring 2010 on *Prunus* in southern France at three sites:
220 near Tautavel (42°47'38N, 2°41'56E), Grabels (43°39'35N, 3°49'12E), and Bompas
221 (42°43'43N, 2°56'31E). We also used collections obtained in spring 2008 near
222 Torreilles (42°44'29N, 2°59'6E). Each sampling site consisted of a small number of
223 bushes or hedges of *Prunus* and covered a few dozen meters at most.

224 Mature adults were sampled at all sites. Progeny (larvae and young adults of the
225 subsequent generation) were sampled at Tautavel and Grabels. Psyllids that fell
226 from beaten branches onto flat nets were either stored in ethanol and/or frozen.
227 Mating pairs caught on nets were stored in separate tubes.

228 Purification of DNA followed Peccoud et al. (2013). Abdomens of mature females
229 were softened in 70% ethanol for spermatophore extraction. Spermatophores were
230 identified as glossy white pellets in spermatheca under a stereo microscope and

231 transferred separately to DNA purification wells. To minimize between-species DNA
232 contamination risk, each batch of dissections, DNA purifications and amplifications
233 of spermatophore DNA was performed on females of the same species.

234 Each DNA sample was assigned to species *A* or *B* using a single diagnostic PCR of the
235 Internal Transcribed Spacer 2 (*ITS2*) gene, which yields an amplicon of a specific
236 size for a given species (Peccoud et al. 2013). Individuals showing two bands –
237 putative hybrids – were reprocessed through DNA extraction (re-using their
238 carcasses after washing in water) and PCR in order to minimize the risk of DNA
239 contamination being interpreted as hybridization. To identify the maternal species
240 of each confirmed hybrid, we Sanger-sequenced a mitochondrial region
241 encompassing the *COI* gene (sequences are available under Genbank under
242 accession numbers xxx). Supporting text I details purification, primers, the
243 genotyping of hybrids and possible sources of error.

244 *Modeling reproductive isolation in Cacopsylla pruni*

245 The genotype data of spermatophores, progeny and mature adults allowed
246 estimation of RI arising: (1) between colonization of *Prunus* by mature adults and
247 insemination, (2) between insemination and the sampling of progeny on *Prunus*, (3)
248 following the sampling of progeny, overwintering on conifers and return of mature
249 adults on *Prunus*. Indices related to (1), (2) and (3) are given superscripts “pre” (for
250 *pre-insemination*), “prog” (*progeny*) and “mat” (*mature adults*) respectively (Figure
251 1).

252 Gene flow due to inseminations at sampling site i , G_{iXY}^{pre} , reflects the proportion of
253 spermatophores of species Y per female of species X (Table 1). All dissected females
254 were inseminated, thus there was no evidence that female reproductive success
255 differed between species. Accordingly, we used equation 8 to model G_{iXY}^{pre} . At this
256 stage, gene flow before the barrier reflects species proportions among mating adults
257 of each sex, thus:

258
$$E_0[G_{iXY}^{\text{pre}}] = f_{iX} \cdot m_{iY},$$

259 where f_{iX} and m_{iY} are the proportions of species X and species Y among mature
260 females and males of site i , respectively. Thus, equation 8 becomes:

261
$$G_{iXY}^{\text{pre}} = \frac{f_{iX} m_{iY} \beta_{iXY}}{m_{iA} \beta_{iXA} + m_{iB} \beta_{iXB}}.$$

262 At a subsequent barrier, null gene flow is gene flow through the previous barrier.

263 Thus, from equation 5, gene flows at the two postzygotic barriers are:

264
$$G_{iXY}^{\text{prog}} = \frac{G_{iXY}^{\text{pre}} \beta_{iXY}^{\text{prog}}}{\sum_{X \in \{A,B\}} \sum_{Y \in \{A,B\}} (\beta_{iXY}^{\text{prog}} G_{iXY}^{\text{pre}})},$$

265
$$G_{iXY}^{\text{mat}} = \frac{G_{iXY}^{\text{prog}} \beta_{iXY}^{\text{mat}}}{\sum_{X \in \{A,B\}} \sum_{Y \in \{A,B\}} (\beta_{iXY}^{\text{mat}} G_{iXY}^{\text{prog}})}.$$

266 G_{iXY}^{prog} reflects the proportion of XY genotypes among progeny, and G_{iXY}^{mat} reflects the
267 proportion of XY genotypes among returning adults. With these specifications,
268 porosities β^{pre} , β^{prog} and β^{mat} can be estimated from genotype data using Bayesian
269 modelling (below).

270 Equation 11 was used to compute the combined porosities of successive barriers
271 (Figure 1). Thus, we quantified β^{host} and RI^{host} , representing RI on spring host-
272 plants [barriers (1)+(2)]; β^{post} and RI^{post} , representing global post-insemination RI
273 [barriers (2)+(3)]; and β^{total} and RI^{total} for barriers (1)+(2)+(3). The “absolute
274 contribution” of barriers (Ramsey et al. 2003; Sobel and Chen 2014) was quantified
275 as the difference in cumulative RI either side of each barrier (Ramsey et al. 2003).

276 *Spatial heterogeneity in pre-insemination isolation*

277 To determine the degree of spatial heterogeneity in pre-insemination RI, β^{pre} , we
278 incorporated a finite mixture model (FMM) (McLachlan and Peel 2000) as a
279 parsimonious model of hidden spatial structure (Pleydell and Chretien 2010). This
280 FMM allocates each study site to one of $k \in \{1 \dots n\}$ “site groups”, where each site in a
281 group shares identical porosities β^{pre} and n is the number of study sites. This
282 introduces vectors \mathbf{z} , which allocates sites to groups, \mathbf{w} , which weights the
283 importance of groups and $\mathbf{\kappa}$, an indicator vector that activates / disactivates groups
284 (see supporting text II).

285 *Bayesian inference*

286 Bayesian analysis of RI in *C. pruni* required making inference from the posterior
287 distribution:

$$288 f(\beta^{\text{pre}}, \beta^{\text{prog}}, \beta^{\text{mat}}, f, m, \kappa, z, w_k | u, v, x, y, p^{\text{obs}}, p^{\text{Mis}}),$$

289 with new terms defined below. Uninformative priors were adopted for all
290 parameters.

291 Likelihood functions for species frequencies among sexes (\mathbf{m} and \mathbf{f}) were obtained
292 assuming the number of individuals of each species among sampled males (\mathbf{u}) and
293 females (\mathbf{v}) follow multinomial distributions with probabilities \mathbf{m} and \mathbf{f} respectively.

294 At barrier (1), the likelihood was evaluated using genotype data for inseminated
295 females (\mathbf{x}) and spermatophores (\mathbf{y}). The numbers of species A and B
296 spermatophores extracted from a female of species X were assumed to follow a
297 multinomial distribution with probabilities $G_{XA} / (G_{XA} + G_{XB})$ and $G_{XB} / (G_{XA} + G_{XB})$,
298 respectively. This assumes independent inseminations – indeed males of the related
299 species *C. pyricola* inseminate one spermatophore per female (Burts and Fisher
300 1967; Krysan 1990).

301 At barrier (2), counts of the four genotypes among progeny, \mathbf{p}^{obs} , were assumed to
302 follow a multinomial distribution with probabilities \mathbf{G}^{prog} . The likelihood also
303 accounted for two hybrids (from Tautavel) of unknown maternal ancestry, \mathbf{p}^{Mis} , that
304 failed to amplify at the mitochondrial region (see supporting text I and II).

305 The likelihood at barrier (3) was derived similarly to that of barrier (2), from
306 genotype data \mathbf{u} and \mathbf{v} , neglecting possible between-year differences in genotype
307 frequency. Further model details, and a glossary defining all variables, are provided
308 in supporting text II.

309 The posterior distribution was sampled using Markov chain Monte Carlo (MCMC)
310 (Gelman et al. 1995). Site-group activation indicators, $\boldsymbol{\kappa}$, were sampled using
311 Reversible Jump MCMC (Green 1995). The model and MCMC algorithm were written
312 and executed in NIMBLE 6.10 (de Valpine et al. 2017) within R 3.4.1 (R Development

313 Core Team 2017). One hundred MCMC chains of 600,000 iterations were run, the
314 first 100,000 iterations were removed as burn-in and samples were saved each 50
315 iterations. Concatenated output (10^6 samples in total) was analyzed using R package
316 CODA. Source code and data are available at
317 https://bitbucket.org/DRJP/reproductive_isolation_mcmc/

318 Results

319 Table 2 summarizes the genotypic data and shows large differences in species
320 frequencies across sites.
321 We did not model premating isolation as only 46 mating pairs were caught on
322 sampling nets, all at Tautavel. Thirty-five involved individuals of species *A*, and 11
323 involved individuals of species *B*. No heterospecific pairs were found. Species
324 proportions in mating pairs were indistinguishable from those in mature adults (χ^2
325 = 0.045, $p = 0.84$, 1 d.f.) but differed significantly from those expected under random
326 mating ($\chi^2 = 40.7$, $p < 0.001$, 1 d.f.). The sampling time of 41 mating pairs sampled
327 over the course of a single day showed little difference between species (Mann-
328 Whitney = 244, $p > 0.19$), providing no evidence for differences in timing of mating
329 activities.
330 Most spermatophores (1812 of the 1990 extracted) were successfully genotyped
331 (missing data is discussed in supporting text I). Interspecific inseminations were
332 detected at all sites (Table 1) and involved 1.38% of genotyped spermatophores.
333 This indicates strong but incomplete pre-insemination isolation (RI^{pre} , Figure 2A).

334 The proportion of MCMC samples in which RI^{pre} differed between sites was ~ 0.001 ,
335 providing only negligible evidence for between-site variation. In terms of
336 asymmetry, RI^{pre} was stronger in $A \times B$ crosses than in the opposite direction, Δ^{pre}
337 being positive (Figure 2G). Other barriers showed little evidence for asymmetry, as
338 posterior distributions of directional RI indices for reciprocal crosses largely
339 overlapped (Δ not shown).

340 Results support positive post-insemination isolation against $B \times A$ hybrids of the
341 progeny (RI_{BA}^{prog} , Figure 2B), meaning these hybrids were less frequent than
342 expected from cross-inseminations. The absence of hybrid genotypes in mature
343 adults (Table 2) rendered RI^{mat} positive for $B \times A$ hybrids (Figure 2C), indicating
344 mortality between emigration from spring hosts and return to these hosts the
345 subsequent year. For the opposite cross direction, uncertainty was large, due to the
346 strong isolation against $A \times B$ insemination and subsequent low expected frequency
347 of $A \times B$ hybrids.

348 The combinations of these successive reproductive barriers led to strong RI^{host} , RI^{post}
349 and essentially complete overall RI (Figure 2D-F). Pre-insemination barriers
350 contributed by far the most to overall RI, as shown by the high absolute contribution
351 of AC^{pre} (Figure 2H).

352 Discussion

353 *Benefits and assumptions of the approach*

354 This work introduces the notion of barrier porosities, which represent contingency-
355 independent probabilities of gene flow, to facilitate RI estimation. Our formulations
356 extend the current RI quantification framework (Sobel and Chen 2014) in several
357 ways.

358 First, they standardize the construction of RI indices for any type of barrier by
359 modelling null gene flows $E_0[\mathbf{G}]$ and potential gene flows \mathbf{G} (Table 1) at each barrier.

360 In addition, by explicitly considering all four cross-types (within and between
361 species), this approach leads naturally to the construction of directional RI indices
362 (RI_{XY} , equation 2). These indices share the properties of Sobel and Chen's (2014)

363 bidirectional RI and satisfy the growing interest in measuring asymmetry in RI
364 (Lowry et al. 2008; Matsubayashi and Katakura 2009; Sanchez-Guillen et al. ;
365 Yukilevich 2012; Brys et al. 2014). A notable difference with the bidirectional RI
366 indices of Sobel and Chen (2014) and of equation 1 is the asymmetry of RI_{XY} , which
367 varies from -3 to 1, and not between -1 and 1. However, a value of -3 accurately
368 informs that directional gene flow is 300% higher than expected under random
369 mating (1/4). It therefore seems sensible to prioritize linearity with gene flow
370 (barrier porosity) over symmetry of the RI index.

371 Our formulation also simplifies the quantification of cumulative effects of sequential
372 and potentially asymmetrical barriers on RI – it is sufficient to compute a

373 normalized product of sequential porosities estimated separately (equation 11).
374 This reduces the difficulty of formulating RI over sequential barriers, where exigent
375 checking for correctness in respect to a particular combination of barriers is
376 typically required (Sobel and Chen 2014). Because barrier porosities are
377 probabilities, and are designed to be contingency-independent, they can be
378 combined (multiplied) for any sequence of barriers studied by any method ranging
379 from field surveys to laboratory experiments (so long as phenotypes controlling RI
380 are not significantly affected by test conditions).
381 Finally, because our porosity-centered specification permits comparison of
382 modelled and observed gene flow, it readily accommodates Bayesian inference and
383 hence credibility intervals for RI-related indices. The potential of Bayesian
384 modelling is demonstrated here with a finite mixture model designed to detect
385 spatial heterogeneity in RI. These developments can help identify local factors
386 conditioning RI. In particular, were RI to vary according to species frequencies, one
387 may question the two main assumptions underlying frequency-independent RI
388 indices: (i) hybrid survival rates are unaffected by genotype frequencies in the
389 progeny (implied by equation 4), neglecting possible effects of competition on
390 hybrids, and (ii) the risk of hybridization per interspecific encounter is stable
391 (equation 7). While we could not evaluate the former assumption due to uneven
392 sampling of progeny and the scarcity of hybrids, the latter is discussed in the next
393 section.

394 *Intensity and contribution of reproductive barriers*

395 The Bayesian model used to analyze genotypic data from *C. pruni* populations
396 demonstrated high, asymmetrical pre-insemination isolation (**RI^{pre}**) with little
397 evidence for between-site variation, and positive post-insemination isolation (**RI^{prog}**
398 and **RI^{mat}**) against *B*×*A* crosses (Figure 2). The combination of these barriers results
399 in essentially complete RI in both directions.

400 Pre-insemination isolation is dominated by premating isolation, given the absence
401 of heterospecific mating pairs among the 46 collected. Conspecific mate preference
402 could be mediated by olfaction (Soroker et al. 2004; Wenninger et al. 2008; Guedot
403 et al. 2009) and/or acoustic signals (Percy et al. 2006; Tishechkin 2007; Wenninger
404 et al. 2009), both of which contribute to species recognition and mate attraction in
405 other psyllid species. Mechanical isolation (Sota and Kubota 1998; Holwell et al.
406 2010) appears unlikely, as variation in male genitalia morphology could not be
407 detected by optical and electron microscopy (N. Sauvion, unpublished). The same
408 can be said for temporal isolation [reviewed in Taylor and Friesen (2017)], as the
409 timing of mating did not significantly differ between species according to mating
410 pairs caught within the course of a day. At larger temporal scales, synchrony
411 between reproductive cycles is supported by the similar species proportions across
412 larval stages at Tautavel ($\chi^2 = 2.0556$, $p > 0.35$, 2 d.f.).

413 We detected that pre-insemination RI significantly differs according to the direction
414 of crosses (Figure 2A,G), suggesting that *B* females and/or *A* males are on average
415 less discriminatory than their allospecific counterparts in respect to species

416 recognition. Asymmetric pre-zygotic isolation is frequently observed in mate
417 preference assays (Jaenike et al. 2006; Rafferty and Boughman 2006; Takami et al.
418 2007; Dopman et al. 2010; Raychoudhury et al. 2010; Merrill et al. 2011; Veen et al.
419 2011; Sanchez-Guillen et al. 2012), but is only rarely measured in the field (Bournez
420 et al. 2015). In comparison to laboratory studies, the asymmetry we observed
421 involves much higher levels of RI (Figure 2A). This suggests that asymmetry in pre-
422 zygotic isolation can persist late in the speciation process, as does prezygotic RI in
423 general (e.g., Coyne and Orr 1997; Mallet et al. 2007; Merrill et al. 2011), and/or that
424 premating RI can be higher in the field than in the laboratory (Jennings and Etges
425 2010).

426 Interestingly, we found no convincing evidence that pre-insemination RI varied
427 between the four sampling sites, despite large differences in relative species
428 frequencies (Table 1). Hence, the assumption of a stable hybridization risk per
429 interspecific encounter (implied by equation 7) is not called into question. Keeping
430 in mind that our ability to challenge this assumption is limited by the number of
431 sampling sites, our observations inform us on the bases of incomplete prezygotic RI
432 in *C. pruni*. A stable risk of mating per interspecific encounter may indicate a certain
433 degree of conspecific mate preference that is both relatively insensitive to site-
434 specific factors, and similar among individuals of the same species and sex (e.g.,
435 Merrill et al. 2011). The hypothesis of between-individual variation in mate
436 preference, potentially due to polymorphism at underlying loci, would less
437 parsimoniously explain incomplete RI. Indeed, it would not explain why cross-
438 inseminations appear more frequent at species *A*-rich sites, unless the less

439 discriminatory individuals essentially occurred among males of species *A*. This
440 hypothesis thus also requires that mate choice be mostly exercised by males. Mate-
441 choice experiments would help to evaluate these hypotheses.

442 The predominant contribution of prezygotic barriers to overall RI (Figure 2H)
443 naturally follows from their early occurrence in the species life-cycle and has been
444 reported in various sympatric species (Ramsey et al. 2003; Malausa et al. 2005; Kay
445 and Husband 2006; Lowry et al. 2008; Matsubayashi and Katakura 2009; Sanchez-
446 Guillen et al. 2012). This predominance does not indicate that post-insemination
447 barriers contributed little to the divergence of *A* and *B* species. These barriers may
448 have reinforced premating isolation [see Servedio and Noor (2003); Coyne and Orr
449 (2004) for a review on reinforcement] and have certainly permitted genetic
450 divergence between *C. pruni* species (Sauvion et al. 2007; Peccoud et al. 2013) in the
451 face of cross-insemination. Some barriers (**RI^{prog}**, Figure 1) operate between
452 insemination and progeny growth, at least for *B*×*A* crosses (Figure 2B), and others
453 (**RI^{mat}**, Figure 2C) affect survival of grown hybrids up to their return on *Prunus*
454 shrubs. In terms of causes, **RI^{prog}** possibly reflects low sperm efficacy in allospecific
455 females (e.g., Matute 2010) and/or reduced hybrid survival up to sampling.

456 Although the scarcity of hybridization in *C. pruni* limits the precision of certain
457 estimates, our case study illustrates how the proposed framework provides
458 estimates of reproductive barriers at an arbitrary number of sampling points
459 through the species life cycle. Future models could incorporate refinements such as
460 independent development and/or survival rates for each sex and developmental

461 stage, more sophisticated models of spatio-temporal variation, or other sources of
462 prior information.

463 **Conflict of interest**

464 The authors declare no conflict of interest.

465

466 Tables

467 **Table 1.** Formulations for null (in the absence of RI at a studied barrier) gene flow
468 due to females of species X and males of species Y ($E_0[G_{XY}]$), and for potential gene
469 flow following modification by the barrier (G_{XY}).

source of RI	$E_0[G_{XY}]$	G_{XY}
ecological niche difference	$f_X \times m_Y^†$	p.‡ of encounters involving females of species X and males of species Y (a) p. of encounters between sexually active mates involving females of species X and males of species Y (b)
allochrony	(a)§	p. of mating pairs involving females of species X and males of species Y (c)
conspecific mate preference	(b)	$f_X \times$ (average proportion of species Y sperm per female of species X) (d)
mechanical incompatibility	(c)	p. of zygotes from females of species X and males of species Y (e)
gametic incompatibility	(d)	p. of progeny from females of species X and males of species Y (f)
early hybrid mortality	(e)	p. of older progeny from females of species X and males of species Y
late hybrid mortality	(f)	

470
471 Barriers are shown in their order of appearance in the life cycle. †f_X : proportion of
472 species X in females; m_Y : proportion of species Y in males. $^‡p.$: proportion. § Letters in
473 parentheses correspond to formulations proposed for G_{XY} in the rightmost column.

474 Formulations may be amended depending of the biological model, for instance
475 individuals may be interpreted as gametes (e.g., pollen for males), and proportion of
476 encounters can be estimated from range or habitat overlap.

477

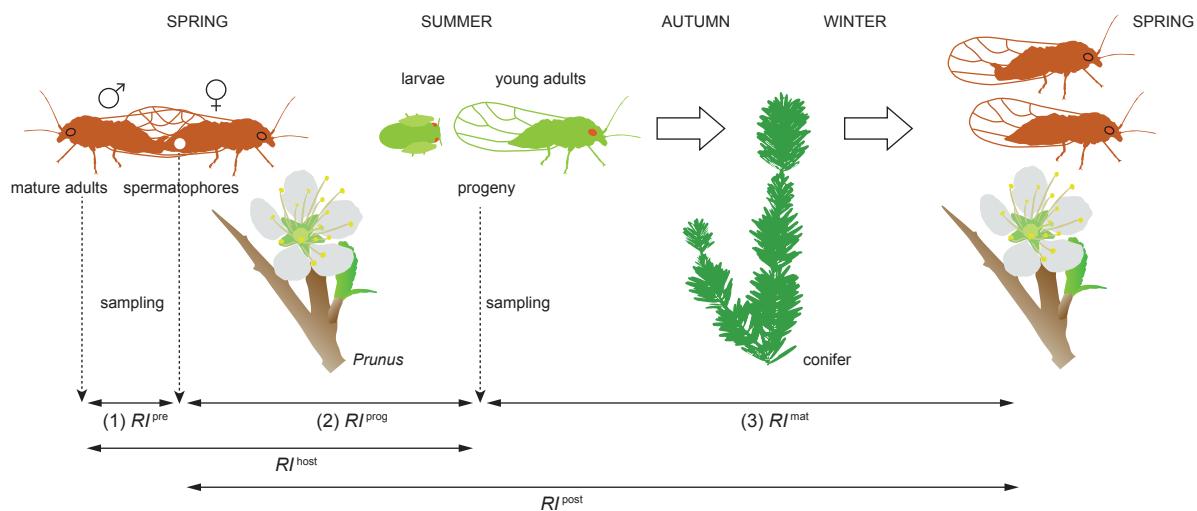
478 **Table 2.** Genotype data (assignment to *C. pruni* species *A* or *B*) from mature adults,
479 spermatophores and progeny at the four sampled sites.

	Tautavel		Grabels		Bompas		Torreilles	
	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>
Females	307	87	127	41	30	31	60	146
Males	211	57	43	18	27	23	26	64
Intra-inseminated females	194	76	110	38	30	30	53	141
Cross-inseminated females	1	11	1	3	0	1	0	5
Spermatophores in <i>A</i> females	826	1	249	1	80	0	83	0
Spermatophores in <i>B</i> females	14	202	3	65	1	70	5	212
Progeny from <i>A</i> mothers	1935	3	382	0				
Progeny from <i>B</i> mothers	10	424	2	142				

480

481 For logistic reasons, not all genotyped females were dissected. “Cross-inseminated”
482 refers to females carrying at least one allospecific spermatophore. For the progeny
483 column headings, *A* or *B*, indicate the paternal species.

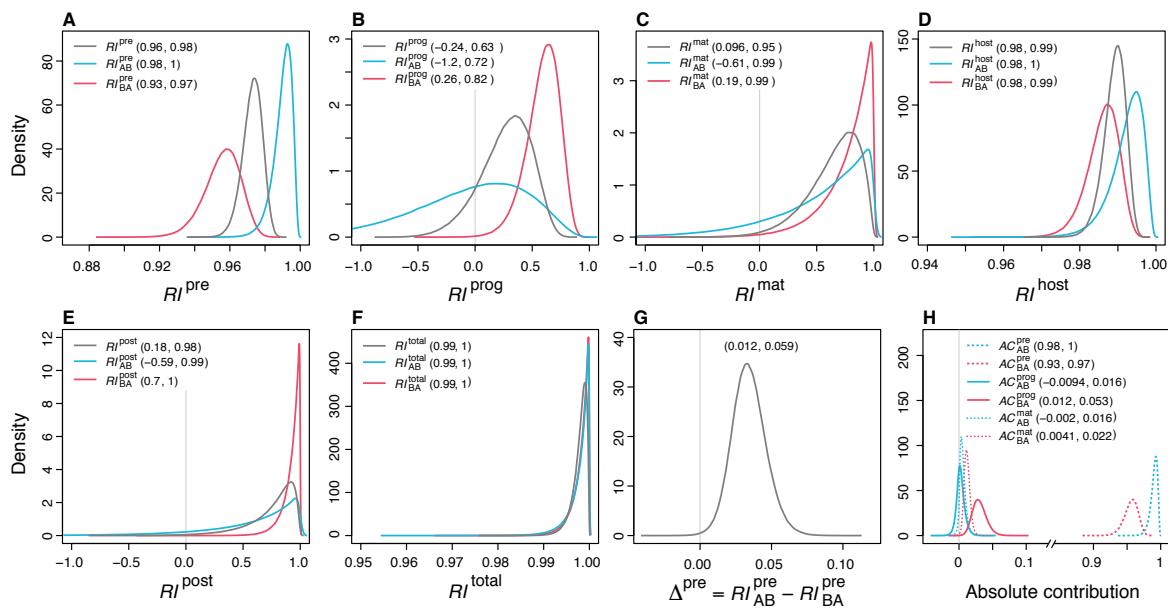
484 **Figures**



485

486 **Figure 1.** Life cycle of *Cacopsylla pruni* and sampling used to estimate reproductive
487 isolation (RI) between its cryptic species at various barriers, or combinations of
488 barriers. Barriers are shown as horizontal arrows and their effects are estimated
489 with RI indices defined in the main text.

490



491

492 **Figure 2.** Posterior probability distributions of reproductive isolation (RI) between
493 *Cacopsylla pruni* species measured at three reproductive barriers (panels A, B, C)
494 and their combinations (panels D, E, F); (G) asymmetry in pre-insemination RI; (H)
495 absolute contributions of reproductive barriers to overall RI. Ninety-five percent
496 credibility intervals are shown in parentheses. See Figure 1 for a representation of
497 the different forms of RI measured in *C. pruni*.

498

499 **References**

500 Beaumont, M. A. 2010. Approximate Bayesian Computation in Evolution and
501 Ecology. *Annu. Rev. Ecol. Evol. Syst.* 41:379-406.

502 Bournez, L., N. Cangi, R. Lancelot, D. R. J. Pleydell, F. Stachurski, J. Bouyer, D.
503 Martinez, T. Lefrancois, L. Neves, and J. Pradel. 2015. Parapatric distribution
504 and sexual competition between two tick species, *Amblyomma variegatum*
505 and *A. hebraeum* (Acari, Ixodidae), in Mozambique. *Parasites & Vectors*
506 8:504.

507 Brys, R., A. Vanden Broeck, J. Mergeay, and H. Jacquemyn. 2014. The contribution of
508 mating system variation to reproductive isolation in two closely related
509 *Centaurium* species (gentianaceae) with a generalized flower morphology.
510 *Evolution* 68:1281-1293.

511 Burts, E. C., and W. R. Fisher. 1967. Mating behavior, egg production and egg fertility
512 in the pear psylla. *J. Econ. Entomol.* 60:1297-1300.

513 Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, R. F. C. Cajas, W. Diao,
514 M. E. Maan, S. Paolucci, F. J. Weissing, L. van de Zande, A. Hoikkala, E.
515 Geuverink, J. Jennings, M. Kankare, K. E. Knott, V. I. Tyukmaeva, C.
516 Zoumadakis, M. G. Ritchie, D. Barker, E. Immonen, M. Kirkpatrick, M. Noor, C.
517 Macias Garcia, T. Schmitt, and M. Schilthuizen. 2012. What do we need to
518 know about speciation? *Trends Ecol. Evol.* 27:27-39.

519 Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecol. Lett.*
520 8:2-14.

521 Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited.
522 Evolution 51:295-303.

523 Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer, Sunderland, USA.

524 Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle. 2009. Accounting for
525 uncertainty in ecological analysis: the strengths and limitations of
526 hierarchical statistical modeling. Ecol. Appl. 19:553-570.

527 de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R.
528 Bodik. 2017. Programming With Models: Writing Statistical Algorithms for
529 General Model Structures With NIMBLE. Journal of Computational and
530 Graphical Statistics 26:403-413.

531 Dopman, E. B., P. S. Robbins, and A. Seaman. 2010. Components of reproductive
532 isolation between north american pheromone strains of the European corn
533 borer. Evolution 64:881-902.

534 Falk, J. J., C. E. Parent, D. Agashe, and D. I. Bolnick. 2012. Drift and selection
535 entwined: asymmetric reproductive isolation in an experimental niche shift.
536 Evol. Ecol. Res. 14:403-423.

537 Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. Bayesian Data Analysis.
538 Chapman and Hall, London.

539 Gompert, Z., E. G. Mandeville, and C. A. Buerkle. 2017. Analysis of Population
540 Genomic Data from Hybrid Zones. Annu. Rev. Ecol., Evol. Syst. 48:207-229.

541 Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and
542 Bayesian model determination. Biometrika 82:711-732.

543 Guedot, C., J. G. Millar, D. R. Horton, and P. J. Landolt. 2009. Identification of a Sex
544 Attractant Pheromone for Male Winterform Pear Psylla, *Cacopsylla pyricola*. J.
545 Chem. Ecol. 35:1437-1447.

546 Hoban, S., G. Bertorelle, and O. E. Gaggiotti. 2012. Computer simulations: tools for
547 population and evolutionary genetics. Nat. Rev. Genet. 13:110-122.

548 Holwell, G. I., C. Winnick, T. Tregenza, and M. E. Herberstein. 2010. Genital shape
549 correlates with sperm transfer success in the praying mantis *Ciulfina klassi*
550 (Insecta: Mantodea). Behav. Ecol. Sociobiol. 64:617-625.

551 Jaenike, J., K. A. Dyer, C. Cornish, and M. S. Minhas. 2006. Asymmetrical
552 reinforcement and *Wolbachia* infection in *Drosophila*. PLoS Biol. 4:1852-
553 1862.

554 Jarausch, W., and B. Jarausch. 2016. A permanent rearing system for *Cacopsylla*
555 *pruni*, the vector of "Candidatus Phytoplasma prunorum". Entomol. Exp. Appl.
556 159:112-116.

557 Jennings, J. H., and W. J. Etges. 2010. Species hybrids in the laboratory but not in
558 nature: a reanalysis of premating isolation between *Drosophila arizonae* and
559 *D. mojavensis*. Evolution 64:587-598.

560 Kaufmann, J., T. L. Lenz, M. Kalbe, M. Milinski, and C. Eizaguirre. 2017. A field
561 reciprocal transplant experiment reveals asymmetric costs of migration
562 between lake and river ecotypes of three-spined sticklebacks (*Gasterosteus*
563 *aculeatus*). J. Evol. Biol. 30:938-950.

564 Kay, K. M., and B. Husband. 2006. Reproductive isolation between two closely
565 related hummingbird-pollinated neotropical gingers. Evolution 60:538-552.

566 Kostyun, J. L., and L. C. Moyle. 2017. Multiple strong postmating and intrinsic
567 postzygotic reproductive barriers isolate florally diverse species of *Jaltomata*
568 (Solanaceae). *Evolution* 71:1556-1571.

569 Krysan, J. L. 1990. Laboratory Study of Mating-Behavior as Related to Diapause in
570 Overwintering *Cacopsylla pyricola* (Homoptera, Psyllidae). *Environ. Entomol.*
571 19:551-557.

572 Lackey, A. C. R., and J. W. Boughman. 2017. Evolution of reproductive isolation in
573 stickleback fish. *Evolution* 71:357-372.

574 Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The
575 strength and genetic basis of reproductive isolating barriers in flowering
576 plants. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* 363:3009-3021.

577 Malausa, T., M. T. Béthenod, A. Bontemps, D. Bourguet, J. M. Cornuet, and S. Ponsard.
578 2005. Assortative mating in sympatric host races of the European corn borer.
579 *Science* 308:258-260.

580 Mallet, J., M. Beltran, W. Neukirchen, and M. Linares. 2007. Natural hybridization in
581 heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.*
582 7.

583 Martin, H., P. Touzet, M. Dufay, C. Gode, E. Schmitt, E. Lahiani, L. F. Delph, and F. Van
584 Rossum. 2017. Lineages of *Silene nutans* developed rapid, strong,
585 asymmetric postzygotic reproductive isolation in allopatry. *Evolution*
586 71:1519-1531.

587 Martin, N. H., and J. H. Willis. 2007. Ecological divergence associated with mating
588 system causes nearly complete reproductive isolation between sympatric
589 *Mimulus* species. *Evolution* 61:68-82.

590 Matsubayashi, K. W., and H. Katakura. 2009. Contribution of multiple isolating
591 barriers to reproductive isolation between a pair of phytophagous ladybird
592 beetles. *Evolution* 63:2563-2580.

593 Matute, D. R. 2010. Reinforcement of Gametic Isolation in *Drosophila*. *PLoS Biol.* 8.

594 McLachlan, G., and D. Peel. 2000. Finite Mixture Models. John Wiley & Sons, Inc,
595 Hoboken, USA.

596 Merrill, R. M., Z. Gompert, L. M. Dembeck, M. R. Kronforst, W. O. McMillan, and C. D.
597 Jiggins. 2011. Mate preference across the speciation continuum in a clade of
598 mimetic butterflies. *Evolution* 65:1489-1500.

599 Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for
600 (incomplete) speciation. *Trends Ecol. Evol.* 24:145-156.

601 Peccoud, J., M. de la Huerta, J. Bonhomme, C. Laurence, Y. Outreman, C. M. Smadja,
602 and J. C. Simon. 2014. Widespread host-dependent hybrid unfitness in the
603 pea aphid species complex. *Evolution* 68:2983-2995.

604 Peccoud, J., G. Labonne, and N. Sauvion. 2013. Molecular Test to Assign Individuals
605 within the *Cacopsylla pruni* Complex. *Plos One* 8:e72454.

606 Percy, D. M., G. S. Taylor, and M. Kennedy. 2006. Psyllid communication: acoustic
607 diversity, mate recognition and phylogenetic signal. *Invertebr. Syst.* 20:431-
608 445.

609 Pleydell, D. R. J., and S. Chretien. 2010. Mixtures of GAMs for habitat suitability
610 analysis with overdispersed presence/absence data. *Comput. Stat. Data Anal.*
611 54:1405-1418.

612 Polacik, M., and M. Reichard. 2011. Asymmetric Reproductive Isolation between
613 Two Sympatric Annual Killifish with Extremely Short Lifespans. *Plos One* 6.

614 Pombi, M., P. Kengne, G. Gimonneau, B. Tene-Fossog, D. Ayala, C. Kamdem, F.
615 Santolamazza, W. M. Guelbeogo, N. Sagnon, V. Petrarca, D. Fontenille, N. J.
616 Besansky, C. Antonio-Nkondjio, R. K. Dabire, A. della Torre, F. Simard, and C.
617 Costantini. 2017. Dissecting functional components of reproductive isolation
618 among closely related sympatric species of the *Anopheles gambiae* complex.
619 *Evolutionary Applications* 10:1102-1120.

620 R Development Core Team. 2017. R: A Language and Environment for Statistical
621 Computing. R Foundation for Statistical Computing, Vienna.

622 Rafferty, N. E., and J. W. Boughman. 2006. Olfactory mate recognition in a sympatric
623 species pair of three-spined sticklebacks. *Behav. Ecol.* 17:965-970.

624 Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive
625 isolation between the monkeyflowers *Mimulus lewisii* and *M-cardinalis*
626 (Phrymaceae). *Evolution* 57:1520-1534.

627 Raychoudhury, R., C. A. Desjardins, J. Buellesbach, D. W. Loehlin, B. K. Grillenberger,
628 L. Beukeboom, T. Schmitt, and J. H. Werren. 2010. Behavioral and genetic
629 characteristics of a new species of *Nasonia*. *Heredity* 104:278-288.

630 Sanchez-Guillen, R. A., M. Wullenreuther, and A. Cordero Rivera. 2012. Strong
631 asymmetry in the relative strengths of prezygotic and postzygotic barriers
632 between two damselfly sister species. *Evolution* 66:690-707.

633 Sauvion, N., O. Lachenaud, G. Genson, J. Y. Rasplus, and G. Labonne. 2007. Are there
634 several biotypes of *Cacopsylla pruni*? *Bull. Insectol.* 60:185-186.

635 Servedio, M. R., and M. A. F. Noor. 2003. The role of reinforcement in speciation:
636 theory and data. *Annu. Rev. Ecol., Evol. Syst.* 34:339-364.

637 Sobel, J. M., and G. F. Chen. 2014. Unification of methods for estimating the strength
638 of reproductive isolation. *Evolution* 68:1511-1522.

639 Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of
640 speciation. *Evolution* 64:295-315.

641 Sobel, J. M., and M. A. Streisfeld. 2015. Strong premating reproductive isolation
642 drives incipient speciation in *Mimulus aurantiacus*. *Evolution* 69:447-461.

643 Soroker, V., S. Talebaev, A. R. Harari, and S. D. Wesley. 2004. The role of chemical
644 cues in host and mate location in the pear psylla *Cacopsylla bidens*
645 (Homoptera : Psyllidae). *J. Insect Behav.* 17:613-626.

646 Sota, T., and K. Kubota. 1998. Genital lock-and-key as a selective agent against
647 hybridization. *Evolution* 52:1507-1513.

648 Takami, Y., N. Nagata, M. Sasabe, and T. Sota. 2007. Asymmetry in reproductive
649 isolation and its effect on directional mitochondrial introgression in the
650 parapatric ground beetles *Carabus yamato* and *C. albrechti*. *Popul. Ecol.*
651 49:337-346.

652 Taylor, R. S., and V. L. Friesen. 2017. The role of allochrony in speciation. *Mol. Ecol.*
653 26:3330-3342.

654 Tishechkin, D. Y. 2007. New data on vibrational communication in psyllids from the
655 families aphalaridae and triozidae (Homoptera, Psylloinea). *Zoologichesky*
656 *Zhurnal* 86:547-553.

657 Veen, T., J. Faulks, R. Rodriguez-Munoz, and T. Tregenza. 2011. Premating
658 Reproductive Barriers between Hybridising Cricket Species Differing in Their
659 Degree of Polyandry. *Plos One* 6.

660 Via, S. 2009. Natural selection in action during speciation. *Proc. Natl. Acad. Sci. USA*
661 106:9939-9946.

662 Wenninger, E. J., D. G. Hall, and R. W. Mankin. 2009. Vibrational Communication
663 Between the Sexes in *Diaphorina citri* (Hemiptera: Psyllidae). *Ann. Entomol.*
664 *Soc. Am.* 102:547-555.

665 Wenninger, E. J., L. L. Stelinski, and D. G. Hall. 2008. Behavioral evidence for a
666 female-produced sex attractant in *Diaphorina citri*. *Entomol. Exp. Appl.*
667 128:450-459.

668 Yukilevich, R. 2012. Asymmetrical patterns of speciation uniquely support
669 reinforcement in *Drosophila*. *Evolution* 66:1430-1446.

670