



HAL
open science

Hybridization in late stages of speciation: Strong but incomplete genome-wide reproductive isolation and ‘large Z-effect’ in a moving hybrid zone

Guilherme Caeiro-Dias, Alan Brelsford, Mariana Meneses-Ribeiro,
Pierre-andré Crochet, Catarina Pinho

► To cite this version:

Guilherme Caeiro-Dias, Alan Brelsford, Mariana Meneses-Ribeiro, Pierre-andré Crochet, Catarina Pinho. Hybridization in late stages of speciation: Strong but incomplete genome-wide reproductive isolation and ‘large Z-effect’ in a moving hybrid zone. *Molecular Ecology*, 2023, 32 (15), pp.4362-4380. 10.1111/mec.17035 . hal-04176054

HAL Id: hal-04176054

<https://hal.science/hal-04176054v1>

Submitted on 2 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

31 Abstract

32

33 In organisms reproducing sexually, speciation occurs when increasing divergence
34 results in pre- or post-zygotic reproductive isolation between lineages. Studies focusing
35 on reproductive isolation origin in early stages of speciation are common; and many
36 rely on genomic scans to infer introgression providing limited information on the
37 genomic architecture of reproductive isolation long-term maintenance. This study
38 analyses a natural hybrid zone between two species in a late stage of speciation. We
39 used ddRADseq genotyping in the contact between *Podarcis bocagei* and *P. carbonelli*
40 to examine admixture extent, analyse hybrid zone stability, and assess genome-wide
41 variation in selection against introgression. We confirmed strong but incomplete
42 reproductive isolation in a bimodal hybrid zone. New findings revealed population
43 genetic structure within *P. carbonelli* in the contact zone; geographic and genomic
44 clines analysis suggested strong selection against gene flow, but a relatively small
45 proportion of the loci can introgress, mostly within the narrow contact zone. However,
46 geographic clines revealed that a few introgressed loci show signs of potential positive
47 selection, particularly into *P. bocagei*. Geographic clines also detected a signal of
48 hybrid zone movement towards *P. bocagei* distribution. Genomic cline analysis
49 revealed heterogeneous patterns of introgression among loci within the syntopy zone,
50 but the majority maintain a strong association with the genomic background of origin.
51 However, incongruences between both cline approaches were found, potentially driven
52 by confounding effects on genomic clines. Last, an important role of the Z chromosome
53 in reproductive isolation is suggested. Importantly, overall patterns of restricted
54 introgression seem to result from numerous strong intrinsic barriers across the
55 genome.

56

57 **Introduction**

58

59 The evolution of reproductive isolation is a gradual process (Templeton, 1992) and
60 hybridization often persists from the initial stages of speciation (Mallet, 2005;
61 Rieseberg, 2009) to more advanced stages (Nosil, Harmon, & Seehausen, 2009).
62 Hybridization and incomplete reproductive isolation may thus result in genetic
63 introgression during most of the duration of the speciation process (Melo-Ferreira,
64 Alves, Freitas, Ferrand, & Boursot, 2009; Mullen, Dopman, & Harrison, 2008; Pinho &
65 Hey, 2010; Tarroso, Pereira, Martínez-Freiría, Godinho, & Brito, 2014). However, as
66 divergence time increases, the build-up of deleterious epistatic interactions between
67 independently evolving loci, also called Bateson-Dobzhansky-Muller incompatibilities
68 (BDMI), impedes gene flow, gradually causing complete reproductive isolation via
69 hybrid unviability or sterility (Bateson, 1909; Dobzhansky, 1937; Gourbiere & Mallet,
70 2010; Matute, Butler, Turissini, & Coyne, 2010; Muller, 1942). Many empirical studies
71 using genomic tools (Ellegren et al., 2012; Gompert et al., 2010; Larson, White, Ross,
72 & Harrison, 2014; Parchman et al., 2013) support the idea that genomes are not
73 homogenous ensembles (Nosil & Feder, 2012; Wu, 2001) and that reproductive
74 isolation affects individual loci differently depending on their linkage to incompatibility
75 genes. Deciphering the nature, number, and distribution of such genetic
76 incompatibilities across the genome remains one of the main challenges in the study
77 of speciation.

78 The build-up of reproductive isolation during divergence can occur through
79 several processes. For example, it can occur due to the accumulation of BDMI starting
80 in 'genomic islands' of reduced gene flow and elevated differentiation (Ellegren et al.,

81 2012; Feder & Nosil, 2010; Turner, Hahn, & Nuzhdin, 2005), becoming more extensive
82 when divergent selection reduces effective gene flow in the surrounding genomic
83 regions, i.e., divergence hitchhiking (Feder & Nosil, 2010; Via, 2012), and may be
84 particularly important for speciation with gene flow if blocks of associated loci that
85 contribute to isolation are co-inherited (Feder, Egan, & Nosil, 2012). Alternative
86 processes that lead to the development of reproductive isolation involve gene
87 incompatibilities arising in many unlinked loci scattered throughout the genome (Feder,
88 Gejji, Yeaman, & Nosil, 2012; Feder & Nosil, 2010) and barrier loci appear later in the
89 speciation process (Dufresnes et al., 2021). Often the evolution of strong reproductive
90 isolation involves multiple barrier effects, interactions among them and the arose of
91 new interactions (Bierne, Welch, Loire, Bonhomme, & David, 2011; Butlin & Smadja,
92 2018; Kulmuni, Butlin, Lucek, Savolainen, & Westram, 2020), but ultimately, as the
93 speciation process proceeds and regardless of the mechanisms involved, the evolution
94 of reproductive isolation is expected to result in identical situations where genomes are
95 too divergent to be compatible.

96 Many studies on the genomics of divergence focus on the early stages of
97 speciation (e.g., Andrew & Rieseberg, 2013; Janoušek et al., 2012; Luttikhuisen, Drent,
98 Peijnenburg, Van der Veer, & Johannesson, 2012; Parchman et al., 2013; Teeter et
99 al., 2010). These early stages are clearly the best to document genomic architectures
100 at the onset of differentiation. However, the factors influencing isolation may vary
101 during different stages of speciation (Ravinet et al., 2017). During the early stages,
102 genetic divergence is low and long-term isolation often rests on pre-mating barriers if
103 divergent populations are sympatric (Egan et al., 2015; van Schooten et al., 2020), but
104 in later stages, the relative contribution of pre- and post-mating barriers often remains
105 an open question. As such, studying genomic mechanisms by which reproductive

106 isolation is maintained (and not only initiated) is equally important. The study of
107 genomic divergence and gene flow in the later stages of speciation is thus crucial to
108 understand which factors influence the genomic architecture of reproductive isolation
109 across the whole process of speciation.

110 Another common question in the genomics of speciation is whether sex
111 chromosomes have a disproportional higher influence on reproductive isolation when
112 compared to autosomes, often formalized as the “large X-effect” and Haldane’s rule.
113 The large X-effect is the higher influence of X (or Z; here referred to as “large Z-effect”)
114 chromosome on hybrid incompatibilities (Coyne & Orr, 2004; Masly & Presgraves,
115 2007; Orr, 1987), as evidenced by restricted introgression on X or Z-linked loci
116 (Payseur, Krenz, & Nachman, 2004; Storchová, Reif, & Nachman, 2010; Tucker, Sage,
117 Warner, Wilson, & Eicher, 1992). Haldane’s rule stems from the observation that
118 fitness reductions are often stronger in hybrids of the heterogametic sex (Haldane,
119 1922), and can cause reduced introgression of loci linked to sex chromosomes and
120 mitochondrial genomes when compared to autosomes (Carling & Brumfield, 2008;
121 Gowen et al., 2014; Walsh, Shriver, Olsen, & Kovach, 2016). Thus, differential patterns
122 of introgression between sex-linked, autosome and mtDNA markers can inform us on
123 the evolutionary forces underlying hybrid incompatibilities.

124 Here we study the genomics of the late stages of speciation by quantifying
125 global and locus-specific patterns of introgression in the only known hybrid zone
126 between two species exhibiting strong but incomplete reproductive isolation: *Podarcis*
127 *bocagei* and *P. carbonelli*. These two species are part of the *Podarcis hispanicus*
128 complex (Iberian and North African wall lizards) and despite their deep divergence (5.6
129 Mya; Kaliontzopoulou, Pinho, Harris, & Carretero, 2011) they still hybridize in a narrow
130 and bimodal hybrid zone (Caeiro-Dias et al., 2021; Pinho, Kaliontzopoulou, Carretero,

131 Harris, & Ferrand, 2009). These species' distributions are allopatric except in this area
132 where both species come into secondary contact. Based on biogeographic and
133 phylogenetic patterns (Pinho, Harris, & Ferrand, 2007; Sá-Sousa, 2001) secondary
134 contact likely happened after post-glacial expansions of both *P. bocagei* and northern
135 populations of *P. carbonelli*. Whether such events of expansion imprinted a genomic
136 signature in the contact zone is still unclear.

137 More specifically, we used ddRADseq genotyping data to i) investigate if the
138 bimodality of the hybrid zone, found in previous studies, can be explained alone by low
139 frequency of hybridization, i.e., if pre-mating barriers between parental species are
140 enough to explain such bimodality; ii) evaluate the extent of introgression and the
141 nature of possible barriers to interspecific gene flow; iii) assess potential effects of
142 population expansion on the stability of the hybrid zone; iv) investigate if loci that show
143 reduced introgression are spread over the genome or grouped in particular regions;
144 and v) if sexual chromosomes have a higher contribution to reproductive isolation than
145 autosomes.

146

147 **Materials and methods**

148

149 *Transect sampling*

150 We collected tissue samples across a 44 km transect including the contact zone
151 between *P. bocagei* and *P. carbonelli* previously studied by Pinho et al. (2009) and
152 Caeiro-Dias et al. (2021, see Figure 1). The contact zone was sampled in a way to
153 include the area of sympatry and adjacent areas to north and south where only *P.*
154 *bocagei* and *P. carbonelli* were found, respectively, to ensure that the complete contact
155 zone was sampled. Here we collected 193 individuals across 2.3 km. We also sampled

156 *P. carbonelli* in three sites south of ESP (SIL, ESM and TOR) and *P. bocagei* in four
157 sites to the north (AGU, FRA, MAD and POR). Sampling details are provided in Figure
158 1, Supplemental Information File 1 (Methods) and Supplemental Information File 2.

159

160 *Extraction of DNA, RAD sequencing, and SNP discovery*

161 Genomic DNA was extracted using the EasySpin® Genomic DNA Tissue Kit (Citomed,
162 Odivelas, Portugal) following the supplier's protocol. The quality and quantity of
163 extracted DNA was evaluated on a 2% agarose gel and a QUBIT 2.0 fluorimeter (Life
164 Technologies, Grand Island, NY, USA).

165 We obtained RAD sequencing data using modifications to protocols
166 from Parchman *et al.* (2012), Peterson *et al.* (2012) and Purcell *et al.* (2014). The
167 complete protocol was described by Brelsford *et al.* (2016) and summarized in
168 Supplemental Information File 1 (Methods). A single library containing all samples was
169 sequenced on eight Illumina® (San Diego, CA, USA) HiSeq 2000 lanes in the
170 Lausanne Genetic Technology Facility (Lausanne, Switzerland), with single-end
171 100 bp reads. Four individuals were included in triplicate to assess sequencing errors.

172 We demultiplexed the raw reads using the process_radtags module of Stacks
173 version 1.3 (Catchen, 2013) allowing one mismatch for barcodes, to remove low-quality
174 reads, reads with uncalled bases and reads that failed the Illumina 'chastity' filter.
175 Adapters were trimmed from demultiplexed reads using a custom shell script. Reads
176 were then aligned in each individual *de novo* to assemble loci using the denovo_map.pl
177 wrap up program. A *de novo* assembly was chosen over the possibility of an assembly
178 guided by a reference genome since the *P. muralis* genome used latter in this study
179 was not yet available at this stage of the work. Furthermore, we discovered enough
180 SNPs using a *de novo* assembly for the type of analysis employed in this study. Each

181 locus was assembled with a minimum depth of sequencing coverage of one and
182 allowing two mismatches between stacks within individuals and two between
183 individuals. We then used VCFtools version 0.1.15 (Danecek *et al.*, 2011) to filter the
184 resulting variants. Loci with depth coverage <8 , with missing data $>20\%$ or with alleles
185 with minimum frequency <0.05 were removed. To avoid possible paralogs, we
186 removed loci containing more than five SNPs and with more than 70% heterozygosity.
187 We selected one SNP per locus by choosing the SNPs maximizing frequency
188 differences between species with a custom Python script available at
189 [https://github.com/catpinho/filter_RADseq_data/blob/master/removes_loc_multipleSN](https://github.com/catpinho/filter_RADseq_data/blob/master/removes_loc_multipleSNPs%2Bleaves1SNP_highfreq.py)
190 [Ps%2Bleaves1SNP_highfreq.py](https://github.com/catpinho/filter_RADseq_data/blob/master/removes_loc_multipleSNPs%2Bleaves1SNP_highfreq.py). This final dataset (complete dataset) was then used
191 to generate two additional datasets, prior to apply a missing data filter by individual.
192 One for loci with allele frequencies higher than 0.8 in reference population containing
193 only parental individuals of one species and lower than 0.2 in the reference population
194 of the other species ("80/20" dataset). The other dataset includes loci with fixed
195 differences between parental populations of *P. bocagei* and *P. carbonell* (diagnostic
196 dataset). The sites POR or TOR were used as the reference populations for *P. bocagei*
197 and *P. carbonelli*, respectively, to produce both datasets. Individuals with missing data
198 $>35\%$ were removed from all datasets.

199

200 *Population structure across the transect*

201 We used Admixture version 1.3.0 (Alexander & Lange, 2011; Alexander, Novembre, &
202 Lange, 2009) to evaluate the proportion of individual genome ancestry (Q) from each
203 parental species on the complete dataset (the proportion of *P. carbonelli* ancestry, Q_C ,
204 is used in this study to describe the results; the proportion of *P. bocagei* ancestry is Q_B
205 $= 1 - Q_C$). Admixture was run until the log-likelihood estimate increased less than 10^{-5}

206 between iterations. We first ran Admixture for $K=2$ because we were mainly interested
207 in individual admixture between the two species. We also extended this analysis from
208 $K=2$ to $K=6$ to detect possible sub-structure within species. The best K was estimated
209 with a five-fold cross-validation error procedure, implemented in Admixture where the
210 best value of K exhibits the lowest cross-validation error. The distributions of Q_c scores
211 in 10 classes of equal size (i.e., classes of 10%) for the contact zone are also reported.

212 Genomic variability among individuals was visualised by performing a principal
213 component analysis (PCA) on the complete dataset with R version 4.0.3 (R Core Team
214 2020) using the adegenet R package version 2.0.1 (Jombart, 2008; Jombart & Ahmed,
215 2011). PCA requires that the dataset do not contain any missing data and thus the
216 missing data values had to be replaced. Instead of using the commonly applied “mean
217 method” to substitute the missing data (where missing data are replaced irrespective
218 of the sample group, here the species), which increases similarity of samples when
219 missing data are numerous and thus mimics the signal of introgression, we used the
220 Breiman’s regression random forest algorithm (Breiman, 2001) implemented in R
221 package randomForest version 4.6-14 (Liaw & Wiener, 2002). The values of our
222 missing data in each of the three datasets were predicted from 500 independently
223 constructed regression trees and 50 bootstrap iterations with default bootstrap sample
224 size.

225 To distinguish putative F1 hybrids in the hybrid zone from other hybrid classes
226 (e.g., F2, backcrosses) and to evaluate the proportion of contemporary hybridization
227 events, we used a triangle plot representing the relationship between interspecific
228 heterozygosity (H_e) and the hybrid index (HI) as described in Fitzpatrick (2012).
229 Further details are provided in Supplemental Information File 1 (Methods). Because in
230 the literature the term “backcross” (BC) is used to refer to both first-generation BC and

231 later generation BC, we would like to clarify that throughout the article we use
232 “backcross” in a general way to refer to both cases, otherwise each specific case will
233 be indicated. We used the diagnostic dataset (1241 loci) to estimate the H_e and an H_I
234 ($H_{I_{DIAG-INT}}$) for each individual in the two populations with signs of admixture (ESP and
235 SIL; see Results), using the R package Introgress version 1.2.3 (Gompert & Buerkle
236 2010), setting *P. bocagei* parental individuals to have an $H_{I_{DIAG-INT}}$ of 0 and *P. carbonelli*
237 parental individuals to 1. The two populations in the extremes of the transect (POR and
238 TOR) were used as the parental populations.

239

240 *Mitochondrial genotyping*

241 The mtDNA lineage of each sample was identified by sequencing a partial fragment of
242 the 12S ribosomal RNA region (12S). Amplification and sequencing were performed
243 as described in Pinho et al. (2006) using the primers published in Kocher et al. (1989).
244 We used individuals from all populations sampled except MAD. Identification was
245 performed by comparing the target sequences with our extensive database of
246 reference sequences using tree building approaches like in Kaliontzopoulou et al.
247 (2011). The identification of mtDNA lineage together with the sex identification of
248 potential F1 and F2 hybrids and first-generation BCs may offer some clues about how
249 F1s, F2s and first-generation BCs were formed, although the number of samples in
250 these hybrid classes was very low (see Results), precluding any statistical analyses.
251 Furthermore, mtDNA lineage was fit a to geographic cline model (*Geographic cline*
252 *analyses* section) to compare with nuclear clines.

253

254 *Simulations of genotypic composition within the contact zone*

255 The low number of backcrosses detected (eight individuals) in our data compared to
256 the number of F1 hybrids (three individuals) suggested poor fitness of F1 hybrids.
257 However, our sample sizes of various hybridization classes were small, so random
258 sampling effects were difficult to exclude. To test whether the observed relative
259 proportion of backcrosses and F1 hybrids in the contact zone can be obtained by
260 random sampling effects without counter-selection of F1 hybrids, we simulated 100
261 replicates of various scenarios combining “random mating” (RM; no restriction to
262 mating between any individual) and “non-random mating” (NRM) with four levels of
263 migration (0, 20%, 30% and 40% of individuals replaced per generation). In NRM
264 scenarios the rate of heterospecific mating was set to the observed proportion of
265 putative F1s to the total number of observed parental individuals in each generation.
266 Once admixed individuals were formed there were no restrictions to gene flow between
267 them and parental or other admixed individuals. A detailed description of simulations
268 is provided in Supporting File 1.

269 Specifically, we tested if immigration and assortative mating between parental
270 individuals in the contact zone are enough to explain the observed proportions of F1
271 hybrids and backcrosses without lower survival, lower fecundity, or lower mating
272 success of F1 hybrids. Each distribution for each simulated scenario (combination of
273 RM and NRM with the 0, 0.05, 0.2, 0.3 and 0.4 of immigrants) was compared to the
274 real distribution using a series of pairwise Fisher’s exact tests and employing the
275 Bonferroni correction for multiple comparisons computed with R package *rcompanion*
276 version 2.0.0 (Mangiafico 2018). Last, we compared the average distribution for each
277 scenario, i.e., the average of each Q score class of 10%, to the observed distribution
278 with similar Fisher’s exact tests. We then tested for each scenario if the ratios of the
279 F1s to backcrosses (F1/BC, where backcrosses are all admixed individuals except

280 F1s) were distinct from the observed data. When the proportion of simulations with
281 higher ratios than the observed data was higher than 0.05, we considered that there
282 was no significant difference between the observed data and the distribution generated
283 by the simulations. All simulation scripts available at
284 http://github.com/catpinho/admixture_simulation.

285

286 *Geographic cline analyses*

287 To perform analyses of geographic clines, we used the R package HZAR
288 (Derryberry, Derryberry, Maley, & Brumfield, 2014). This program provides functions
289 to fit allele frequency data to equilibrium geographic cline models (Barton & Gale, 1993;
290 Gay, Crochet, Bell, & Lenormand, 2008; Szymura & Barton, 1986, 1991) using the
291 Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm. To each sampling
292 location was assigned a distance along the transect that corresponds to the shortest
293 distance to POR population along the coastline (i.e., avoiding a straight line over the
294 ocean). Allele frequencies at each SNP locus included in the “80/20” dataset were fitted
295 to 16 equilibrium geographic cline models using HZAR version 3.0.3 (Derryberry et al.,
296 2014). We further used the Q_C values from the Admixture analysis as hybrid indices
297 from the complete (HI) and diagnostic datasets ($HI_{DIAG-ADMIX}$) and also the mtDNA locus
298 (*P. bocagei* or *P. carbonelli* maternal lineage) to fit to the same 16 models. Details
299 about estimated parameters are provided in the Supplemental Information File 1
300 (Methods). We performed three independent runs of 1,000,000 MCMC iterations, a
301 burn-in of 100,000 iterations and sampling every 10 iterations for each model and
302 checked for convergence. For each SNP locus, HI , $HI_{DIAG-ADMIX}$, and mtDNA, the model
303 with the lowest AIC score was selected as the best-fitting model. We then estimated
304 the CI as the region delimited by the maximum and minimum values of the 95%

305 credible cline region (Derryberry et al. 2014). To test if each SNP locus significantly
306 differed from the genome average, we compared its centre (c) and width (w) with the
307 same parameters for the *HI* cline (c_{HI} and w_{HI}), which we took as a sort of a “genome-
308 wide average”. We considered each parameter for individual SNP clines to exhibit
309 significant differences from the *HI* cline if their 95% CI did not overlap. We also
310 searched for outlier loci based on the locus-specific estimates of c and w . Finally, to
311 test if the number of SNP loci with centres shifted to north or south were similar or
312 significantly different, we performed a binomial test. We did not try to estimate selection
313 and dispersal using the approach of (Barton & Gale, 1993) as this assumes random
314 mating within the hybrid zone, an assumption that seems unlikely given the low number
315 of F1 hybrids in our sample (see below).

316

317 *Genomic clines analyses*

318 We quantified patterns of genomic introgression in admixed populations using the BGC
319 software version 1.0 (Gompert & Buerkle, 2012). We performed preliminary analysis
320 including the total number of loci and individuals, but to decrease the confounding
321 effect of population structure (see Discussion), we report the results including only *P.*
322 *bocagei* and *P. carbonelli* individuals occurring in syntopy, using the complete dataset.
323 BGC uses a Bayesian genomic cline model (Gompert & Buerkle, 2011, 2012; Gompert
324 et al., 2012a) to estimate ancestries of individuals summarized as a hybrid index
325 (HI_{BGC}) and to quantify genome-wide variation in introgression among admixed
326 populations. The model includes parameter α and β that describe, for every locus, the
327 transition in allele frequency from the genomic background of origin (ancestry) to the
328 genomic background of the other species, along the gradient of admixed genotype
329 classes (i.e., HI_{BGC}) between the two parental species (Gompert & Buerkle, 2011).

330 Details about each parameter are provided in Supplemental Information File 1
331 (Methods). Despite the potential interest in estimating these parameters, simulation
332 studies have shown that deviations in α relative to genomic average can be explained
333 by underdominance when dispersal is low, directional selection or BDMI, whereas β
334 can be affected by underdominance when dispersal is high or by BDMI (Gompert &
335 Buerkle, 2011). However, BDMI have limited effect on genomic cline parameters
336 (Gompert & Buerkle, 2011). As such, multiple selection types can result in similar
337 patterns of α and β making its interpretation difficult and hence the interpretation of
338 genomic clines should be made with caution.

339 *A priori* knowledge of pure parental individuals is required to estimate the
340 proportion of the ancestry attributed to each parental species. We used the populations
341 POR and MAD as *P. bocagei* parental individuals and ESM and TOR as *P. carbonelli*,
342 while using ESP as the test (admixed) population. With the complete dataset we ran
343 three independent MCMC chains for 300,000 steps each and recorded samples from
344 the posterior distribution every 25th step following a 200,000-step burn-in to estimate
345 marginal posterior probability distributions for H_{BGC} and cline parameters α and β . We
346 combined the output of the two chains after inspecting the MCMC output to assess
347 convergence to the stationary distribution.

348 The significance of individual α and β values was assessed by checking whether
349 their 95% CI included zero or not. These estimates provide a locus-specific view on
350 the process of introgression between species, but they do not compare how genomic
351 clines differ between loci. Therefore, in addition we searched for outlier loci on the basis
352 of the genome-wide distribution of locus-specific estimates of α and β values by
353 determining the cline parameter quantiles as recommended by Gompert & Buerkle
354 (2011, 2012).

355

356 *Comparison between geographic and genomic clines*

357 To understand whether geographic and genomic cline parameters show
358 concordance in the patterns of reproductive isolation, we tested if the 2300 loci
359 common to “80-20” and complete datasets are correlated, using Person’s correlation
360 tests implemented in R. Moreover, we checked whether the same loci showed
361 increased patterns of introgression isolation to compare the consistency of the results
362 obtained with both approaches.

363

364 *Distribution of cline parameters across the genome and “Large Z effect”*

365 To identify the position of each SNP on the genome we aligned each of the 6905
366 RADtag sequences from the complete dataset, including the 2300 “80/20” SNPs, with
367 the *Podarcis muralis* genome (Andrade et al., 2019). *Podarcis muralis* is the closest
368 relative to the Iberian and North African clade of *Podarcis* (the *Podarcis hispanicus*
369 complex) to which *P. bocagei* and *P. carbonelli* belong; they diverged around 10 million
370 years ago (Yang et al., 2021). The alignment was performed using the blastn algorithm
371 (Altschul et al., 1997) included in the BLAST+ package version 2.9.0 (Camacho et al.,
372 2009).

373 If many loci more or less equally distributed across the genome contribute to
374 species barrier, it is expected that levels of introgression are positively correlated with
375 recombination rate (Aeschbacher, Selby, Willis, & Coop, 2017; Martin, Davey, Salazar,
376 & Jiggins, 2019). Here we used chromosome length as a proxy to recombination rate
377 as chromosome length is negatively correlated with average recombination rate. We
378 tested if cline parameters are correlated with chromosome size using Person’s

379 correlation tests on the “80-20” (for geographic cline parameters) and complete (for
380 genomic cline parameters) datasets using the SNPs that aligned to *P. muralis* genome.

381 Furthermore, we tested if the distribution of values from each parameter used
382 to describe introgression, from both geographic (c and w) or genomic (α and β) cline
383 analyses, were similar between chromosomes and mtDNA, using non-parametric
384 Kruskal-Wallis tests. If the test was significant for any of the parameters, we then
385 performed pairwise Wilcoxon tests among chromosomes using a Bonferroni correction
386 for multiple comparisons. These tests were performed with the R package *stats* version
387 4.0.3 (R Core Team, 2020). In a scenario of a “large Z-effect”, we expect introgression
388 on Z-linked loci to be significantly lower than on autosome loci.

389

390 **Results**

391 *SNP and mtDNA datasets*

392 The final dataset consisted of 6905 SNPs (complete dataset) with an average
393 depth of coverage of 28 and median of 26.8 (min = 12.5; max = 85.8) and 329
394 individuals with an average depth of coverage of 29 and median of 28.8 (min = 15.6;
395 max = 48.6). Using this dataset, we assessed the replicability to be on average 99.87%
396 for the tested samples. Additionally, we filtered the complete dataset to obtain two
397 additional datasets, the “80/20” dataset containing of 2300 loci with allele frequencies
398 higher than 0.8 in reference populations of one species and lower than 0.2 in the other,
399 and the diagnostic dataset containing 1241 loci with fixed frequencies in each parental
400 population. Finally, for mtDNA dataset, we successfully sequenced the 12S fragment
401 from 336 individuals. Details about the final sample size for each dataset is provided
402 in Supplemental Information File 1 (Table S1).

403

404 *Population structure across the transect*

405 As expected from previous works (see Caeiro-Dias et al., 2021; Pinho et al.,
406 2009), the current contact zone between *P. bocagei* and *P. carbonelli* was located in
407 ESP population and was mostly made of parental individuals (strongly bimodal hybrid
408 zone; Figure 2): Admixture identified 147 individuals as parental from which 114 had
409 $Q_C \geq 0.9$ and 33 $Q_C \leq 0.1$, including 66 with Q_C of one and 21 with Q_C of zero (Figure
410 2a). The remaining 14 individuals captured in the contact zone showed signs of large
411 proportion of admixture ($0.1 < Q_C < 0.9$). Actual syntopy was restricted to a ~450 meters
412 wide stripe in ESP including 115 samples. Based on multilocus assignment results,
413 this length was delimited by the most northern *P. carbonelli* with $Q_C \geq 0.9$ and the
414 most southern *P. bocagei* with $Q_C \leq 0.1$. These proportions are commonly used as
415 thresholds to distinguish parental from admixed genotypes. Within the syntopy area,
416 102 individuals were assigned to parental genotypes, 90 with a genomic background
417 assigned to *P. carbonelli* and 12 to *P. bocagei*, including 52 with $Q_C = 1$ and three with
418 $Q_C = 0$ (Figure 2b). The binomial test ($p < 0.001$) shows that the number of samples in
419 this area is biased towards *P. carbonelli*, suggesting a higher local density of this
420 species since nothing indicated higher capturability. The other 13 individuals in the
421 syntopy area showed signs of recent mixed ancestry. Summary results for Q_C scores
422 per population are presented in Supplemental Information File 1, Table S1.
423 Accordingly, the first PCA axis (PC1) explains 57.1% of the variation in the data and
424 clearly distinguishes *P. bocagei* and *P. carbonelli* populations (Figure 2c). PC2
425 explains 1.4% of the variation and corresponded to genetic structure within *P.*
426 *carbonelli*, separating the ESP population from all other *P. carbonelli* populations. In
427 fact, when we increased K with Admixture, K=3 (Figure 2a) was the best scenario to
428 explain the variation in the SNP data. In concordance with the PCA, *P. carbonelli*

429 individuals from ESP were mostly assigned to a group apart from the other *P. carbonelli*
430 populations showing sub-structuring within *P. carbonelli*.

431 The plot of *He* against *Hi* (Figure 2d) showed results concordant with Admixture
432 and PCA. One individual was close to the expectations for first-generation BC and
433 three others were close to the expectations for F1 hybrids, although *He* was lower than
434 1. False homozygote call due to PCR duplicates is a known source of bias across
435 RADseq protocols (for discussion on this see Andrews et al., 2014; Puritz et al., 2014)
436 and would reduce *He* even in F1 hybrids. While the single-end reads ddRADseq
437 protocol used in this study did not allow us to detect PCR duplicates (Andrews et al.,
438 2014; Puritz et al., 2014), we performed four independent PCR for each sample, to
439 decrease the proportion of PCR duplicates in the final library. Moreover, we were able
440 to detect two F1s with interspecific *He* = 1 in a contact zone between *P. carbonelli* and
441 *P. lusitanicus* using the same protocol (Caeiro-Dias et al., 2021). An alternative
442 explanation is that those three potential F1s were a product of crosses between
443 parental individuals from different species with residual levels of admixture, which is a
444 plausible scenario since in the contact zone we observed a continuum of admixed
445 proportions. While our results can be explained by one of these scenarios or by a
446 combination of both, the observed *He* and *Hi* values for these three individuals were
447 closer to F1s than any other hybrid class and thus they were considered as F1 hybrids.
448 Those three individuals were adult males, all with *P. bocagei* mtDNA. The potential
449 first-generation BC was also identified as an adult male with mtDNA assigned to *P.*
450 *carbonelli*.

451 An admixture signal was also evident in SIL population south to ESP since all
452 individuals in SIL were identified as *P. carbonelli* on the basis of morphology and
453 mtDNA but exhibited 0.02 to 0.05 of admixture from *P. bocagei* in SNP data (for $K=2$,

454 Fig. 2a); SIL was the only population away from the contact zone with evident signs of
455 admixture.

456

457 *Simulations of genotypic composition within the contact zone*

458 Simulations suggested that most combinations of RM or NRM (in our model it
459 was restricted to parental individuals) with various migration rates were unable to
460 produce distributions of genotype classes matching our observed distribution. Only for
461 NRM with 40% of parental genotypes migrating into the syntopy zone the simulations
462 were not significantly different from our data (the corrected p -value = 0.461 was higher
463 than 0.05; Supplemental Information File 1, Table S2). Without higher levels of
464 migration, with either RM or NRM, F1 hybrid genotypes were no longer observed in a
465 few generations.

466 The proportions of F1s relative to backcrosses for most scenarios with RM and
467 NRM is lower than in observed data (Figure 3). Note that the relative proportions of
468 F1s was the same in the observed data from the syntopy area and simulated data. The
469 observed proportion was not significantly different in only three scenarios (RM0.3,
470 RM0.4 and NRM0.4; Supplemental Information File 1, Table S3). Such results
471 suggested less backcrosses than expected in the observed data given the number of
472 F1 hybrids and hence a reduced fitness of F1 hybrids.

473

474 *Geographic clines*

475 The fixed/right geographic cline model (p_{min}/p_{max} fixed to zero and one with right
476 tail fitted, i.e., estimated tail south of the hybrid zone towards *P. carbonelli*) was the
477 best-fit model for the *Hl* cline, with $c_{HI} = 13.15\text{km}$ (CI = 11.98–14.04km) south of POR
478 (ESP; Figure 4) and $w_{HI} = 0.54\text{km}$ (CI = 0.1–3.18km) indicating a very narrow hybrid

479 zone. Tail parameters δ and τ were 0.43km and 0.238, respectively. The best-fitted
480 models varied across the 2300 loci (see Supplemental Information File 1, Table S4)
481 but most of the geographic clines for each SNP locus were highly concordant in both
482 c and w (Figure 4a and Supplemental Information File 1, Table S5) and not significantly
483 different from the HI estimates (Table 1). Only 3.2% of SNPs have a cline centre that
484 is shifted north and 2.1% shifted south. The binomial test revealed that these
485 proportions are significantly different ($p < 0.001$) from random expectations. From those
486 SNPs with c significantly shifted, we detected six outlier loci shifted to north but none
487 to south (Figure 4a). Also, just 8.1% of the loci had significantly wider clines than the
488 genomic average but none had a narrower cline. Detecting statistically significant lower
489 w was not possible since w_{HI} was already very narrow and so was the lower margin of
490 the confidence interval ($w_{HI} = 0.54\text{km}$, $CI = 0.1\text{--}3.18\text{km}$). The regression model
491 estimated for c versus w suggested that there a significant linear correlation between
492 these two parameters, i.e., clines with c shifted to north tended to have larger w (Figure
493 4a). Another aspect was that while for 59.4% of the loci the best fitted model did not
494 include tail parameter estimations, for 28% the model included a right tail (to south),
495 which was also reflected in the none/right model for the HI . From these loci, 98% were
496 mapped to autosomes. On the other hand, the proportion of loci fitted to a model with
497 left tail (to north) was only 3%, and from these, 95.7% were autosomal loci. For the
498 $HI_{DIAG-ADMIX}$ the best-fitted model was the none/mirror (p_{min}/p_{max} not estimated and
499 symmetric tails), with $c_{HI-DIAG-ADMIX} = 13.55\text{km}$ ($CI = 13.18\text{--}13.92\text{km}$) south of POR and
500 $w_{HI-DIAG-ADMIX} = 1.99\text{km}$ ($CI = 0.74\text{--}3.74\text{km}$). Thus, both HI estimation did not produce
501 significantly different cline centres. Finally, for the mtDNA cline the none/none (tails
502 and p_{min}/p_{max} not estimated) was the best-fitted model, with $c_{mtDNA} = 14.12\text{km}$ ($CI =$
503 $13.06\text{--}14.14\text{km}$) and $w_{mtDNA} = 0.1\text{km}$ ($CI = 0.01\text{--}3.74\text{km}$). The mtDNA cline centre had

504 overlapping CIs with both HI estimates and hence the nuclear and the mtDNA clines
505 were not statistically different.

506

507 *Genomic clines*

508 Preliminary results based on the total number of individuals sampled in the
509 contact zone revealed the same overall pattern of α and β distribution. The main
510 difference is that the parameter β shows more extreme positive values which conforms
511 to the expectations in the presence of marked population structure. We thus focus our
512 analysis in the syntropy area. The genomic clines within the syntropy area showed high
513 heterogeneity across the 6905 loci ($-8.03 < \alpha < 5.43$; $-18.70 < \beta < 24.97$) with many loci
514 significantly deviating from the expectations given the HI estimated from the complete
515 dataset (Table 1, Figure 4b and Figure 5). We detected a significant excess of *P.*
516 *bocagei* ancestry (lower bound of 95% CI for $\alpha > 0$) in 15.2% of the loci and an excess
517 of *P. carbonelli* ancestry (upper bound of 95% CI for $\alpha < 0$) in 32.3%. These proportions
518 were significantly different from chance expectations (binomial test $p < 0.001$). A total of
519 52.5% of the loci did not present an excess of ancestry for any species given the HI
520 (their 95% CI for α includes 0). We also detected 18.7% of loci with significantly steeper
521 changes in allele frequencies (lower bound of 95% CI for $\beta > 0$), 16.6% with
522 significantly shallower changes in allele frequencies (upper bound of 95% CI for $\beta < 0$)
523 and 64.7% following the expectations given the HI (95% CI for β includes 0). From the
524 1053 loci with excess of *P. bocagei* ancestry, 19.5% follow the expectations given the
525 HI_{BGC} , 8.8% exhibit steep changes in allele frequencies and 71.7% showed signs of
526 increased introgression. For the 2227 loci with excess of *P. carbonelli* ancestry, 48%
527 follow the expectations given the HI_{BGC} , 50.4% of the loci showed restricted

528 introgression and only 1.7% had shallower changes in allele frequencies. No outlier
529 loci were detected for the observed distributions of α and β .

530

531 *Comparison between geographic and genomic clines*

532 By comparing geographic and genomic clines from the 2300 loci common to
533 “80-20” and complete datasets, we detected a negative but weak correlation between
534 c and α (Figure 5a) and between w and β (Figure 5b). However, many loci have cline
535 parameters that contrast with the significant correlations. For example, a considerable
536 proportion of loci, despite increasing introgression in the syntopy zone (significantly
537 negative β ; 17.3%) or no differences from the genomic average (β non-significant;
538 29.6%), did not show significant w differences from the w_{HI} . Moreover, 18.9% of loci
539 show significant negative β , i.e., increased introgression when compared to the rest of
540 the genome, and the majority of those loci (17.4% of the total) had narrow w similar to
541 w_{HI} . Only the remaining 1.6% of loci showed congruent patterns of significant
542 introgression.

543 Additionally, we found inconsistent patterns of loci exhibiting increased
544 introgression in the alternative genomic background between both approaches. While
545 geographic clines showed 2.6% of loci with c shifted to north and w wider than w_{HI} , only
546 0.1% of the loci showed excess *P. carbonelli* ancestry with higher introgression, and
547 neither set had loci in common. Conversely, 0.5% presented c shifted to south and
548 wide clines, whereas 0.4% had excess *P. bocagei* ancestry and higher introgression
549 but only 0.1% (three loci) were common to both sets. These results show that both
550 approaches identify different sets of loci with alleles that might be under positive
551 selection in the foreigner genomic background.

552

553 *Distribution of cline parameters across the genome*

554 We successfully aligned on the *P. muralis* genome 6027 RADtags (out of the
555 6905 from the complete dataset) and 2101 (out of the 2300 present in the “80/20”
556 dataset), representing a density of 1 marker per 236Kb and 1 marker per 678Kb,
557 respectively.

558 We did not identify significant correlations between any cline parameter and
559 chromosome length. When testing using geographic cline parameters, for c $R^2 = 0.01$
560 and p -value = 0.68; for w $R^2 = 0.21$ and p -value = 0.05. When using genomic cline
561 parameters, for α $R^2 = 0.05$ and p -value = 0.36; for β $R^2 = 0.01$ and p -value = 0.57.
562 Thus, levels of introgression do not seem to be correlated with recombination rate.

563 The distributions of geographic and genomic cline parameters and the densities
564 of values across each chromosome are represented in Figure 6. These plots,
565 particularly the violin plots depicting density values (Figure 6b-e), show striking
566 differences between the Z chromosome and the autosomes. Specifically, in the Z
567 chromosome c and w are less variable, w is noticeably lower, α is skewed towards *P.*
568 *carbonelli* ancestry (negative values), and β has a more marked bimodal distribution
569 than the autosomes. The difference in the geographic cline parameter c was small but
570 significant between the Z chromosome (mean = 13.84km) and all autosomes (mean =
571 13.37km; Table 2 and Supplemental Information File 1, Table S6), while the distribution
572 of w was significantly different between the Z chromosome and all autosomes but one,
573 the chromosome 18 (Table 2 and Supplemental Information File 1, Table S7).
574 However, it is worth to note the lack of outliers in chromosome 18 (Figure 6c) and that
575 the distribution of w in this autosome was not statistically different from the other
576 autosomes. For the genomic cline parameter α we found significant differences only
577 between Z and three other chromosomes (Table 2 and Supplemental Information File

578 1, Table S8), while for β all comparisons between the Z chromosome and all
579 autosomes were significant, except with chromosomes 8 and 18 (Table 2 and
580 Supplemental Information File 1, Table S9). Overall, these results suggest lower
581 introgression in Z-linked loci.

582

583 **Discussion**

584

585 *Genomic composition in the hybrid zone*

586 As shown by previous works (Caeiro-Dias et al., 2021; Pinho et al., 2009), the
587 hybrid zone between *P. bocagei* and *P. carbonelli* is very narrow and made mostly of
588 parental genotypes. Nearly half of the individuals in the syntopy area did not show any
589 evidence of admixture, although the proportions of parental individuals are largely
590 biased towards *P. carbonelli*. However, the other half of the individuals presented a
591 gradient of admixed proportions including three F1 hybrids, recent and later generation
592 backcrosses, and residual levels of admixture. This continuum of assignment
593 proportions across admixed genotypes shows that the two species hybridize and
594 produce fertile offspring resulting in recurrent backcrosses over many generations, as
595 previously demonstrated in Caeiro-Dias et al. (2020). Moreover, we did not find any
596 mito-nuclear discordance outside of the contact zone.

597 Outside of the contact zone, substantial signs of admixture between both
598 species were almost undetected, except in SIL population. The similar low frequencies
599 of *P. bocagei* assignment across all individuals in SIL suggest that after gene flow
600 between this population and ESP in the past, these populations may now be isolated,
601 or gene flow has been highly restricted for several generations.

602 Within *P. carbonelli*, the individuals from ESP exhibit a genetic composition
603 different from the other *P. carbonelli* populations across the transect, independently
604 from *P. bocagei* admixture. This genetic structure within *P. carbonelli* suggests
605 differentiation between ESP and southern populations. A comprehensive examination
606 of population structure within the range of *P. carbonelli* would be needed to interpret
607 this result.

608

609 *Reduced hybrid fitness*

610 In our simulation study, most of the simulated distributions of admixture classes
611 do not match the data observed in the syntopy area. The simulated scenarios were
612 similar to the observed data only when we implemented high migration levels (i.e., the
613 replacement of 40% of the individuals in the contact zone by migrants in each
614 generation). Given that the syntopy range is small (about 450 m) this scenario is
615 possible but extreme. In other scenarios with less dispersal or without assortative
616 mating (irrespective of dispersal), the effects of selection are necessary to explain our
617 results. When compared to the simulated data, the frequency of BC relative to the
618 number of F1s in the contact zone was low, suggesting a combination of lower
619 fecundity of F1 and/or lower survival of BC individuals. Pinho et al. (2009) failed to
620 detect abnormalities of the gonads or reduced number of spermatozoa in admixed
621 individuals of this contact zone but we don't have direct estimates of fecundity or
622 survival so we cannot conclude on the selective processes acting on hybrids there.
623 Furthermore, although the three F1s are males with *P. bocagei* mtDNA, the BC male
624 with *P. carbonelli* do not offer any clue suggesting that one sex may be less fit than the
625 other since it was necessary a female F1 with *P. carbonelli* mtDNA to originate the BC.
626 We recognize that assortative mating can indeed be stronger than what was

627 implemented in our model and may play a more important role than what our results
628 suggest. In conclusion, strong pre- or post-mating barriers are necessary to explain the
629 low number of F1 and BC we see in the contact zone, but we cannot yet decipher their
630 relative contributions, although they both likely contribute to reproductive isolation in
631 this system. To fully evaluate relative the importance of pre-mating isolation, low
632 survival or low fertility of F1 and low survival of BC on reproductive isolation in this
633 system, we would need larger sample sizes of various genotype classes combined with
634 simulations incorporating more complex and realistic assortative mating scenarios
635 (e.g., Irwin, 2020).

636

637 *Strong selection against introgression maintains a restricted hybrid zone*

638 The geographic cline analyses and the estimation of selection in the centre of
639 the contact zone provided further evidence of strong selection against gene flow
640 between the two species. The geographic clines show a high proportion of loci (91.9%)
641 with steep clines, indicating the presence of strong barriers to gene flow spread over
642 the genome between populations of both species apart from the contact zone (Mallet
643 et al., 1990), and most of the clines (89.7%) had approximately the same cline centre
644 and width. These patterns are coherent with a tension zone model where multiple
645 unlinked loci form concordant steep clines at the centre of the hybrid zone (Barton,
646 1983; Barton & Gale, 1993; Barton & Hewitt, 1985). In a stage where substantial
647 barriers already exist, coupling of barrier effects can be achieved by processes like
648 spatial movement of clines (Bierne et al., 2011) or cycles of population expansion and
649 contraction (Hewitt, 1989), leading to concordant steep clines in the late stages of
650 speciation. Both spatial movement of clines and population fluctuations can explain

651 current genetic variation across Iberian *Podarcis* (Pinho, Harris, & Ferrand, 2007; Sá-
652 Sousa, 2001).

653

654 *Differential introgression patterns across the genome detected within the contact zone*

655 In contrast to the overall concordance among loci detected with geographic cline
656 analysis, genomic clines show larger heterogeneity of introgression patterns among
657 loci within the syntopy area, particularly for β . However, positive values of β may be
658 inflated in the presence of population structure within the hybrid zone (Gompert &
659 Buerkle, 2011, 2012; Gompert et al., 2012a). *Podarcis carbonelli* individuals used as
660 reference (TOR) and those from the contact zone (ESP) were clustered in different
661 groups with PCA and Admixture and thus true β values may be lower than our
662 estimations. Nevertheless, genomic cline results indicate that most of the loci do not
663 introgress more or less than what would be expected given the Hl_{BGC} , which can be
664 explained because most loci maintain a strong linkage with the other loci on the
665 genomic background of origin (Gompert et al., 2012b), preventing intraspecific gene
666 flow. Epistatic interactions such as BDMI are considered to be the most common
667 intrinsic barriers to gene flow (Coyne & Orr, 2004; Qvarnström & Bailey, 2009) and in
668 the early stages of speciation are expected to generate patterns of differential
669 introgression across the genome because of stochasticity in the accumulation of such
670 incompatibilities (Brennan, Hiscock, & Abbott, 2014; Fishman, Kelly, Morgan, & Willis,
671 2001; Scascitelli et al., 2010). However, in the late stages of speciation enough BDMI
672 should have build-up across the genome to confer strong reproductive isolation. Across
673 *P. hispanicus* complex allopatric divergence is likely the best explanation for climatic
674 niche divergence (Caeiro-Dias et al., 2018) and thus the patterns of reproductive
675 isolation across the genomes of *P. bocagei* and *P. carbonelli* may be related with

676 divergence in allopatry, where a large portion of the genome may have become
677 differentiated through selective or neutral processes (Fierst & Hansen, 2010; Gavrilets,
678 Hai, & Vose, 1998; Gompert et al., 2012b; Lynch & Force, 2000). The evolution of
679 current strong reproductive isolation likely involved different processes, interactions
680 between processes and potential new interactions arising along the evolution of
681 speciation (Kulmuni et al., 2020). More explicitly, intricated interactions between
682 genetic architecture of intrinsic barriers to gene flow, populations history, and selected
683 and neutral variation, might be contributing to the strong reproductive isolation (Bierne
684 et al., 2011; Butlin & Smadja, 2018; Kulmuni et al., 2020) but disentangling the effect
685 of each process is not easy. Whilst genomic clines are of limited usefulness to detect
686 pairwise epistatic loci (Gompert & Buerkle, 2011; Gompert & Buerkle, 2009), our
687 results show that in the late stages of speciation, most loci do not show signs of
688 increased introgression compared to genomic average, consistent with the hypothesis
689 that the genomic architecture of reproductive isolation is determined by many widely
690 dispersed genomic regions.

691

692 *Comparison of geographic and genomic clines*

693 A direct comparison of geographic and genomic cline parameters between loci
694 common to both analyses detected significant but weak negative correlations.
695 Significant correlations between c and α were driven mostly because some loci with
696 excess of *P. carbonelli* ancestry have very high values of c (shifted to north). When
697 comparing w and β , significant results were obtained mostly because almost half of the
698 loci (44.9%) had steep clines (not significantly different from w_{HI}) and decreased
699 introgression in the syntopy zone (significantly positive β), showing a concordance in
700 the lack of introgression between both cline analyses. However, weak correlations can

701 be explained because regardless of the variation observed in the genomic cline
702 parameters, the geographic cline parameters were consistent for most loci denoting a
703 strong reproductive isolation and a lack of substantial introgression outside of the
704 contact zone. A particularly interesting result was that 17.3% of the loci showed wide
705 genomic clines but narrow geographic width raising the question whether extrinsic
706 selection might play a role constraining interspecific gene flow outside of the contact
707 zone. While it is a plausible explanation because the two species exhibit significantly
708 different climatic niches (Caeiro-Dias et al., 2018) it need further research.

709 Additionally, we found inconsistent patterns of loci potentially under positive
710 selection between both approaches. Geographic and genomic clines identified,
711 respectively, 2.6% and 0.1% of loci with *P. carbonelli* alleles with increased
712 introgressing in *P. bocagei* genomic background, but none of the loci are common to
713 both sets. Conversely, the same approaches identified, respectively, 0.5% and 0.4%
714 of loci with *P. bocagei* alleles with increased introgressing in *P. carbonelli* genomic
715 background but only 0.1% were loci identified by both methods. Loci identified by
716 genomic clines in the syntopy zone exhibiting increased introgression are not expected
717 to necessarily introgress away from the contact zone, but at least the loci with signs of
718 increased introgression in populations apart from the contact zone should show
719 consistent patterns of introgression with both approaches, regardless of what selective
720 forces are contributing to the geographic patterns of introgression detected. While
721 these results could be explained by *P. carbonelli* population structure, as it is expected
722 to potentially inflate the signal of decreased introgression detected with genomic clines
723 within the syntopy zone (Gompert & Buerkle, 2011, 2012; Gompert et al., 2012a), it is
724 not clear in the available literature to date how to undoubtedly interpret such
725 differences. In particular, because genomic cline analyses may be sensitive to many

726 distinct selective and demographic phenomena (Gompert & Buerkle, 2011) we cannot
727 rule out the hypothesis that the differences suggested by both approaches are led by
728 these confounding effects. We thus suggest that additional testing on simulated and
729 empirical datasets using genomic cline approaches are needed to disentangle the
730 causes of inconsistent results using different methods.

731

732 *Genomic signatures of a moving hybrid zone and potential positive selection*
733 *northwards*

734 Results show that several loci introgress asymmetrically. About 28% of
735 geographic clines were fitted to a model with an estimated tail to south while only 3%
736 were fitted to a model with a tail estimated to north. In both cases, the large majority
737 are autosome loci. These results are consistent with a scenario of a moving hybrid
738 zone northwards in the direction of *P. bocagei*, since a moving hybrid zone is expected
739 to leave a tail of clines of unlinked neutral markers resulting in asymmetrical cline
740 variation (Buggs 2007; Teeter *et al.* 2010), essentially driven by autosomes. Although
741 this pattern could alternatively be interpreted as unidirectional introgression away from
742 a static hybrid zone (Moran, 1981), the consistency in the proportions of assignment
743 to *P. bocagei* across SIL individuals suggests that the northward movement of the
744 hybrid zone left *P. bocagei* alleles behind, and currently the frequencies are
745 homogenized. The overall genomic clines pattern is also concordant with this
746 interpretation; 13.7% of the loci with increased *P. bocagei* ancestry and increased
747 introgression is considerably higher than the 4.4% of loci with restricted introgression,
748 indicating that several loci with excess *P. bocagei* ancestry can introgress in the
749 genomic background of *P. carbonelli*. On the other hand, the 1.2% of loci with
750 increased *P. carbonelli* ancestry and increased introgression is much lower than 15.8%

751 with restricted introgression, showing that many loci with increased *P. carbonelli*
752 ancestry do not introgress in the genomic background of *P. bocagei*.

753 Interestingly, a restricted group of loci exhibit an opposite pattern: the 2% of loci
754 have *c* shifted to south on average 1.65km from the *HI* cline centre, while 3.2% of clines
755 with *c* shifted to north on average 5.03km. These proportions are statistically different.
756 Also, clines with centre shifted to north tend to be wider than other clines (both
757 parameters are positively correlated), wider clines than the *HI* cline, and include outlier
758 loci. Loci with *c* shifted from the majority of other clines have been interpreted as being
759 subject to positive selection on one side of the hybrid zone (Baldassarre, White,
760 Karubian, & Webster, 2014; Ballard & Whitlock, 2004; Barton & Hewitt, 1985; Teeter &
761 Payseur, 2008). It is reasonable to hypothesise that loci with *c* shifted to north are good
762 candidates for harbouring genes with *P. carbonelli* alleles under positive selection that
763 confer advantage in the genomic background or habitat of *P. bocagei*. Note, however,
764 that genetic drift in hybrid zones can increase false positives when testing for adaptive
765 introgression (Jofre & Rosenthal, 2021) and affect clines shape (Polechová & Barton,
766 2011). Drift is higher in smaller demes with lower migration rates among demes. In our
767 species, dispersal is probably highly restricted among local populations (small sized
768 terrestrial nonflying vertebrate in fragmented environment), but local populations are
769 apparently quite large, which should prevent strong drift. Nevertheless, we cannot
770 entirely discard at this stage that drift rather than selection generated variation in cline
771 centre among loci. Simulations or direct tests of selection would be necessary to
772 undoubtably rule out the effect of drift.

773

774 *Distribution of cline parameters across the genome and “Large Z effect”*

775 We did not find any correlation between chromosome length and cline
776 parameters, suggesting that barrier loci are distributed across the genome and show
777 little variation in introgression levels. In the late stages of speciation, it is expected that
778 loci are in such strong linkage with the genomic background of origin that even a high
779 recombination rate is not enough to restore linkage in admixed genotypes.

780 The w of loci on the Z chromosome was significantly smaller on average than
781 that of most autosomes (the only exception being chromosome 18). Similar results
782 were obtained for β , that was significantly smaller in the Z than in most of autosomes,
783 except chromosome 8 and 18. Despite non-significant differences between Z and
784 chromosomes 8 and 18, these two chromosomes do not show consistent significant
785 differences when compared to other autosomes. We interpret these results as a
786 consistent signature of lower introgression in the Z chromosome compared to the
787 autosomes. Such differences are expected when the sexual chromosomes have a
788 distinct role in the reproductive isolation compared to autosomes and are concordant
789 with the “large X effect” (or “large Z effect” in the case of the ZW sex determination
790 system) which is a long-standing principle of speciation and has been identified to play
791 a role in introgression patterns and reproductive isolation across many species (e.g.,
792 Dufresnes et al., 2021; Guénet, et al., 1990; Jiggins et al., 2001; Llopart, 2012; Orr,
793 1987; Storchová et al., 2010).

794 However, differences in effective population sizes between autosomes and Z/X
795 chromosomes are also expected to generate steeper clines (higher slope, lower width)
796 for Z/X compared to autosomes as smaller effective size result in stronger effect of drift
797 (Polechová & Barton, 2011) even if there is no difference in selection. Differences in
798 slopes due to drift are difficult to predict, as they depend on selection, dispersal, and
799 effective population size, with strong selection, strong dispersal and large population

800 size resulting in low drift and similar slopes for autosomes and sex chromosomes. In
801 our hybrid zone, we inferred strong selection, and neither dispersal nor population size
802 appear to be restricted. We thus discard drift as the main explanation for the steeper Z
803 clines compared to autosomes.

804 Lasne et al. (2017) reported that clines generated by local adaptation are
805 steeper for loci on the X or Z chromosomes than for loci on autosomes, but in this case
806 the cause of the systematically higher slope for the X/Z is that the X/Z is more
807 responsive to selection than the autosomes are (T. Lenormand and T. Connallon, pers.
808 com.). This effect only applies to primary clines generated by selection of alternative
809 alleles across an environmental change in space and does not affect tension zones
810 maintained by a balance between dispersal and selection against hybrids/intergrades,
811 as the slope of clines in tension zones only depends on selection and dispersal (Barton
812 & Gale 1993), both of which are independent of effective population size.

813 Another possible explanation for steeper Z clines would be female-biased
814 dispersal, as a higher dispersal of the heterogametic sex tend to increase the difference
815 in slopes between Z/X and autosomes. However, in squamates, males tend to disperse
816 more (see Ferchaud et al., 2015 and references therein), including an example in the
817 *Podarcis* genus (Vignoli, Vuerich, & Bologna, 2012). Moreover, mtDNA geographical
818 cline parameters c and w were not statistically different from the nuclear *Hl* cline
819 parameters, suggesting that dispersion is not female-biased. Even if there is no
820 information on sex biases in dispersal in our focal species, we can thus probably
821 discard sex-biased dispersal as an explanation here for the steeper Z clines compared
822 to the neutral expectation relative to the autosomes. Unequal sex-ratio and differences
823 in the variance of reproductive success can also lead to deviations from neutral
824 expectations for Z and autosome cline slope ratio. While sex ratio in our dataset is not

825 different from 1:1 (observed 1 female to 1.2 males; $X^2 = 0.703$; p -value = 0.402),
826 differences in reproductive success are not possible to exclude.

827 Based on our results, steeper Z clines are thus likely the result of “large Z-effect”.
828 Furthermore, we detected a discordance in c between Z and autosomes. The
829 geographical cline centres of Z-linked loci are in general displaced to the south
830 (essentially matching with the area of syntopy) from autosome cline centres (generally
831 matching with the c_{HI} , slightly to the north of the syntopy area). Although this difference
832 is only of a few hundred meters, it is nevertheless significant. In the context of a moving
833 hybrid zone towards north (previously discussed), the displacement of Z cline centres
834 to the south is consistent with the scenario of “large Z effect” where Z cline centres
835 essentially match the centre of the contact zone and do not show signs of introgression.

836

837 *Concluding remarks*

838 The strong bimodality of the hybrid zone between *P. bocagei* and *P. carbonelli* and the
839 patterns of strong reproductive isolation between these two deeply divergent species
840 confirm that these species have reached the late stages of speciation. Accordingly, we
841 found multiple lines of evidence of selection against hybrids. However, even in this late
842 stage of speciation, a relatively small proportion of the loci can still introgress more
843 than the genomic average, but such introgression is mostly asymmetric and suggests
844 the hybrid zone is moving towards *P. bocagei* distribution range. Furthermore, we
845 found evidence that the Z chromosome plays a disproportionately large role in
846 reproductive isolation. Overall, our results support a model of increasingly numerous
847 barriers as speciation proceeds resulting in a genomic architecture of reproductive
848 isolation in the late stages of speciation that relies on cohesion across the genome
849 against interspecific gene flow determined by multiple interacting barriers.

850

851 **Acknowledgments**

852 We thank E. Pérez, M. A. Carretero, N. Sillero, A. Kaliontzopoulou, A. Perera, A. Crotini
853 and A. Salas for helping with sample collection and Thomas Lenormand, Luis-Miguel
854 Chevin, Tim Connallon and Clémentine Lasne for their help with cline analyses.
855 Molecular Ecology's subject editor and two reviewers made highly valuable comments
856 that significantly improved the manuscript. This study benefited from the Montpellier
857 Bioinformatics Biodiversity platform supported by the LabEx CeMEB, an ANR
858 *Investissements d'avenir* program (ANR-10-LABX-04-01). GCD was supported by a
859 PhD grant (SFRH/BD/89750/2012), and CP by an IF contract
860 (IF/01597/2014/CP1256/CT0009), under the *Programa Operacional Potencial*
861 *Humano—Quadro de Referência Estratégico Nacional* funds from the European Social
862 Fund and Portuguese *Ministério da Educação e Ciência* and PAC by the ANR grant
863 ANR-19-CE02-0011 – IntroSpec. Support was also provided by national funds through
864 FCT projects: PTDC/BIA-BEC/102179/2008 – FCOMP-01-0124-FEDER-007062,
865 under FEDER COMPETE funds; PTDC/BIA-EVL/28090/2017 – POCI-01-0145-
866 FEDER-028090 and PTDC/BIA-EVL/30288/2017 – NORTE-01-0145-FEDER-30288
867 co-funded by NORTE2020 through Portugal 2020 and FEDER Funds. Specimens for
868 this study were captured and handled under permit numbers 486 to 490/2014/CAPT
869 by *Instituto da Conservação da Natureza e das Florestas* (ICNF, Portugal).

870

871 **References**

872 Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic
873 inference of the strength and timing of selection against gene flow. *Proceedings*
874 *of the National Academy of Sciences*, 114(27), 7061-7066.

- 875 Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm
876 for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246.
- 877 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of
878 ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. doi:
879 10.1101/gr.094052.109.vidual
- 880 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., &
881 Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of
882 protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.
- 883 Andrade, P., Pinho, C., i de Lanuza, G. P., Afonso, S., Brejcha, J., Rubin, C.-J., ...
884 Bellati, A. (2019). Regulatory changes in pterin and carotenoid genes underlie
885 balanced color polymorphisms in the wall lizard. *Proceedings of the National
886 Academy of Sciences*, *116*(12), 5633–5642.
- 887 Andrew, R. L., & Rieseberg, L. H. (2013). Divergence is focused on few genomic
888 regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*,
889 *67*(9), 2468–2482.
- 890 Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart,
891 G. (2014). Trade-offs and utility of alternative RADseq methods: reply to Puritz et
892 al. *Molecular Ecology*, *23*, 5943-5946.
- 893 Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and
894 morphological analysis of a semipermeable avian hybrid zone suggests
895 asymmetrical introgression of a sexual signal. *Evolution*, *68*(9), 2644–2657. doi:
896 10.1111/evo.12457
- 897 Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of
898 mitochondria. *Molecular Ecology*, *13*(4), 729–744.
- 899 Barton, N. H. (1983). Multilocus clines. *Evolution*, *37*(3), 454–471.

- 900 Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In Harrison, R.
901 G. (Ed.). *Hybrid Zones and the Evolutionary Process*, 13–45. Oxford University
902 Press.
- 903 Barton, Nicholas H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review*
904 *of Ecology and Systematics*, 16(1), 113–148.
- 905 Bateson, W. (1909). Heredity and variation in modern lights. In: Seward, A.C. (Ed.),
906 *Darwin and Modern Science*. Cambridge University Press, Cambridge, 85-101.
- 907 Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- 908 Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage
909 maps of a European tree frog (*Hyla arborea*) identify the sex chromosome
910 without information on offspring sex. *Heredity*, 116(2), 177.
- 911 Brennan, A. C., Hiscock, S. J., & Abbott, R. J. (2014). Interspecific crossing and
912 genetic mapping reveal intrinsic genomic incompatibility between two *Senecio*
913 species that form a hybrid zone on Mount Etna, Sicily. *Heredity*, 113(3), 195.
- 914 Caeiro-Dias, G., Brelsford, A., Kaliontzopoulou, A., Meneses-Ribeiro, M., Crochet, P.-
915 A., & Pinho, C. (2021). Variable levels of introgression between the endangered
916 *Podarcis carbonelli* and highly divergent congeneric species. *Heredity*, 126, 463-
917 476.
- 918 Caeiro-Dias, G., Luís, C., Pinho, C., Crochet, P. A., Sillero, N., & Kaliontzopoulou, A.
919 (2018). Lack of congruence of genetic and niche divergence in *Podarcis*
920 *hispanicus* complex. *Journal of Zoological Systematics and Evolutionary*
921 *Research*. doi: 10.1111/jzs.12219
- 922 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., &
923 Madden, T. L. (2009). BLAST+: architecture and applications. *BMC*
924 *Bioinformatics*, 10(1), 1–9.

- 925 Carling, M. D., & Brumfield, R. T. (2008). Haldane's rule in an avian system: using
926 cline theory and divergence population genetics to test for differential
927 introgression of mitochondrial, autosomal, and sex-linked loci across the
928 *Passerina* bunting hybrid zone. *Evolution*, 62(10), 2600–2615.
- 929 Catchen, J. M. (2013). Stacks: an analysis tool set for population genomics.
930 *Molecular Ecology*, 22(11), 3124–3140. doi: 10.1111/mec.12354.Stacks
- 931 Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA. Sinauer Associates,
932 Inc.
- 933 Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR:
934 hybrid zone analysis using an R software package. *Molecular Ecology*
935 *Resources*, 14(3), 652–663.
- 936 Dobzhansky, T. G. (1937). *Genetics and the Origin of Species* (Vol. 11). Columbia
937 University Press.
- 938 Dufresnes, C., Brelsford, A., Jeffries, D. L., Mazepa, G., Suchan, T., Canestrelli, D.,
939 ... Martínez-Solano, I. (2021). Mass of genes rather than master genes underlie
940 the genomic architecture of amphibian speciation. *Proceedings of the National*
941 *Academy of Sciences*, 118(36).
- 942 Egan, S. P., Ragland, G. J., Assour, L., Powell, T. H. Q., Hood, G. R., Emrich, S., ...
943 Feder, J. L. (2015). Experimental evidence of genome-wide impact of ecological
944 selection during early stages of speciation-with-gene-flow. *Ecology Letters*,
945 18(8), 817–825.
- 946 Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ...
947 Qvarnström, A. (2012). The genomic landscape of species divergence in
948 *Ficedula* flycatchers. *Nature*, 491(7426), 756.
- 949 Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-

- 950 flow. *Trends in Genetics*, 28(7), 342–350.
- 951 Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new
952 mutations under divergence and genome hitchhiking. *Philosophical Transactions*
953 *of the Royal Society B: Biological Sciences*, 367(1587), 461–474.
- 954 Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating
955 genomic islands during ecological speciation. *Evolution: International Journal of*
956 *Organic Evolution*, 64(6), 1729–1747.
- 957 Ferchaud, A., Eudeline, R., Arnal, V., Cheylan, M., Pottier, G., Leblois, R., & Crochet,
958 P. (2015). Congruent signals of population history but radically different patterns
959 of genetic diversity between mitochondrial and nuclear markers in a mountain
960 lizard. *Molecular Ecology*, 24(1), 192–207.
- 961 Fierst, J. L., & Hansen, T. F. (2010). Genetic architecture and postzygotic
962 reproductive isolation: evolution of Bateson–Dobzhansky–Muller incompatibilities
963 in a polygenic model. *Evolution: International Journal of Organic Evolution*,
964 64(3), 675–693.
- 965 Fishman, L., Kelly, A. J., Morgan, E., & Willis, J. H. (2001). A genetic map in the
966 *Mimulus guttatus* species complex reveals transmission ratio distortion due to
967 heterospecific interactions. *Genetics*, 159(4), 1701–1716.
- 968 Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using
969 molecular markers. *BMC Evolutionary Biology*, 12(1), 131.
- 970 Gavrillets, S., Hai, L., & Vose, M. D. (1998). Rapid parapatric speciation on holey
971 adaptive landscapes. *Proceedings of the Royal Society of London B: Biological*
972 *Sciences*, 265(1405), 1483–1489.
- 973 Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on
974 molecular and phenotypic traits in hybrid zones: A window on tension zone

- 975 models. *Evolution*, 62(11), 2789–2806. doi: 10.1111/j.1558-5646.2008.00491.x
- 976 Gompert, Z., & Buerkle, C. A. (2009). A powerful regression-based method for
977 admixture mapping of isolation across the genome of hybrids. *Molecular*
978 *Ecology*, 18(6), 1207–1224. doi: 10.1111/j.1365-294X.2009.04098.x
- 979 Gompert, Z., & Buerkle, C. A. (2010). INTROGRESS: a software package for
980 mapping components of isolation in hybrids. *Molecular Ecology Resources*,
981 10(2), 378–384.
- 982 Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines.
983 *Molecular Ecology*, 20, 2111–2127. doi: 10.1111/1755-0998.12009.x
- 984 Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of
985 genomic clines. *Molecular Ecology Resources*, 12, 1168–1176. doi: doi:
986 10.1111/1755-0998.12009.x
- 987 Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012a). Genomics of isolation in
988 hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
989 367(1587), 439–450. doi: 10.1098/rstb.2011.0196
- 990 Gompert, Z., Forister, M. L., Fordyce, J. A., Nice, C. C., Williamson, R. J., & Buerkle,
991 C. A. (2010). Bayesian analysis of molecular variance in pyrosequences
992 quantifies population genetic structure across the genome of *Lycaeides*
993 butterflies. *Molecular Ecology*, 19(12), 2455–2473. doi: 10.1111/j.1365-
994 294X.2010.04666.x
- 995 Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A.
996 (2012b). Genomic regions with a history of divergent selection affect fitness of
997 hybrids between two butterfly species. *Evolution*, 66(7), 2167–2181.
- 998 Gourbiere, S., & Mallet, J. (2010). Are species real? The shape of the species
999 boundary with exponential failure, reinforcement, and the “missing snowball.”

- 1000 *Evolution*, 64(1), 1–24.
- 1001 Gowen, F. C., Maley, J. M., Cicero, C., Peterson, A. T., Faircloth, B. C., Warr, T. C.,
1002 & McCormack, J. E. (2014). Speciation in Western Scrub-Jays, Haldane’s rule,
1003 and genetic clines in secondary contact. *BMC Evolutionary Biology*, 14(1), 1–15.
- 1004 Guénet, J.-L., Nagamine, C., Simon-Chazottes, D., Montagutelli, X., & Bonhomme, F.
1005 (1990). Hst-3: an X-linked hybrid sterility gene. *Genetics Research*, 56(2–3),
1006 163–165.
- 1007 Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of*
1008 *Genetics*, 12(2), 101–109.
- 1009 Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M.
1010 W., ... Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation
1011 in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m.*
1012 *domesticus*. *Molecular Ecology*, 21(12), 3032–3047.
- 1013 Jiggins, C. D., Linares, M., Naisbit, R. E., Salazar, C., Yang, Z. H., & Mallet, J.
1014 (2001). Sex-linked hybrid sterility in a butterfly. *Evolution*, 55(8), 1631–1638.
- 1015 Jofre, G. I., & Rosenthal, G. G. (2021). A narrow window for geographic cline
1016 analysis using genomic data: Effects of age, drift, and migration on error rates.
1017 *Molecular Ecology Resources*, 21(7), 2278–2287.
- 1018 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic
1019 markers. *Bioinformatics*, 24(11), 1403–1405.
- 1020 Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of
1021 genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071.
- 1022 Kaliontzopoulou, A., Pinho, C., Harris, D. J., & Carretero, M. A. (2011). When cryptic
1023 diversity blurs the picture: A cautionary tale from Iberian and North African
1024 *Podarcis wall* lizards. *Biological Journal of the Linnean Society*, 103(4), 779–800.

- 1025 doi: 10.1111/j.1095-8312.2011.01703.x
- 1026 Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V, Pääbo, S., Villablanca, F.
1027 X., & Wilson, a C. (1989). Dynamics of mitochondrial DNA evolution in animals:
1028 amplification and sequencing with conserved primers. *Proceedings of the*
1029 *National Academy of Sciences of the United States of America*, 86(16), 6196–
1030 6200. doi: 10.1073/pnas.86.16.6196
- 1031 Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the
1032 maintenance of species boundaries. *Molecular Ecology*, 23(7), 1668–1678. doi:
1033 10.1111/mec.12601
- 1034 Lasne, C., Sgrò, C. M., & Connallon, T. (2017). The relative contributions of the X
1035 chromosome and autosomes to local adaptation. *Genetics*, 205(3), 1285–1304.
- 1036 Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R*
1037 *News*, 2(3), 18–22.
- 1038 Llopart, A. (2012). The rapid evolution of X-linked male-biased gene expression and
1039 the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids.
1040 *Molecular Biology and Evolution*, 29(12), 3873–3886.
- 1041 Luttikhuisen, P. C., Drent, J., Peijnenburg, K., Van der Veer, H. W., & Johannesson,
1042 K. (2012). Genetic architecture in a marine hybrid zone: comparing outlier
1043 detection and genomic clines analysis in the bivalve *Macoma balthica*. *Molecular*
1044 *Ecology*, 21(12), 3048–3061.
- 1045 Lynch, M., & Force, A. G. (2000). The origin of interspecific genomic incompatibility
1046 via gene duplication. *The American Naturalist*, 156(6), 590–605.
- 1047 Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and*
1048 *Evolution*, 20(5), 229–237. doi: 10.1016/j.tree.2005.02.010
- 1049 Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M., & Eeley, H. (1990).

- 1050 Estimates of selection and gene flow from measures of cline width and linkage
1051 disequilibrium in *Heliconius* hybrid zones. *Genetics*, 124(4), 921–936.
- 1052 Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate
1053 variation shapes barriers to introgression across butterfly genomes. *PLoS*
1054 *biology*, 17(2), e2006288.
- 1055 Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of
1056 the two rules of speciation in *Drosophila*. *PLoS Biology*, 5(9), e243.
- 1057 Matute, D. R., Butler, I. A., Turissini, D. A., & Coyne, J. A. (2010). A test of the
1058 snowball theory for the rate of evolution of hybrid incompatibilities. *Science*,
1059 329(5998), 1518–1521.
- 1060 Melo-Ferreira, J., Alves, P. C., Freitas, H., Ferrand, N., & Boursot, P. (2009). The
1061 genomic legacy from the extinct *Lepus timidus* to the three hare species of
1062 Iberia: Contrast between mtDNA, sex chromosomes and autosomes. *Molecular*
1063 *Ecology*, 18(12), 2643–2658. doi: 10.1111/j.1365-294X.2009.04221.x
- 1064 Moran, C. (1981). Genetic demarcation of geographical distribution by hybrid zones.
1065 *Proc. Ecol. Soc. Aust*, 11, 67–73.
- 1066 Mullen, S. P., Dopman, E. B., & Harrison, R. G. (2008). Hybrid zone origins, species
1067 boundaries, and the evolution of wing-pattern diversity in a polytypic species
1068 complex of North American admiral butterflies (Nymphalidae: *Limenitis*).
1069 *Evolution*, 62(6), 1400–1417. doi: 10.1111/j.1558-5646.2008.00366.x
- 1070 Muller, H. (1942). Isolating mechanisms, evolution, and temperature. *Biol. Symp.*, 6,
1071 71–125.
- 1072 Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and
1073 consequences. *Philosophical Transactions of the Royal Society B: Biological*
1074 *Sciences*, 367(1587): 332-342.

- 1075 Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for
1076 (incomplete) speciation. *Trends in Ecology & Evolution*, *24*(3), 145–156.
- 1077 Orr, H. A. (1987). Genetics of male and female sterility in hybrids of *Drosophila*
1078 *pseudoobscura* and *D. persimilis*. *Genetics*, *116*(4), 555–563.
- 1079 Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J.
1080 A. C., ... Buerkle, C. A. (2013). The genomic consequences of adaptive
1081 divergence and reproductive isolation between species of manakins. *Molecular*
1082 *Ecology*, *22*(12), 3304–3317. doi: 10.1111/mec.12201
- 1083 Parchman, Thomas L, Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., &
1084 Buerkle, C. (2012). Genome-wide association genetics of an adaptive trait in
1085 lodgepole pine. *Molecular Ecology*, *21*(12), 2991–3005.
- 1086 Payseur, B. A., Krenz, J. G., & Nachman, M. W. (2004). Differential patterns of
1087 introgression across the X chromosome in a hybrid zone between two species of
1088 house mice. *Evolution*, *58*(9), 2064–2078.
- 1089 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012).
1090 Double digest RADseq: an inexpensive method for de novo SNP discovery and
1091 genotyping in model and non-model species. *PLoS One*, *7*(5), e37135.
- 1092 Pinho, C., Kaliontzopoulou, A., Carretero, M. A., Harris, D. J., & Ferrand, N. (2009).
1093 Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and
1094 *Podarcis carbonelli*: Evidence for limited natural hybridization and a bimodal
1095 hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*,
1096 *47*(4), 368–377. doi: 10.1111/j.1439-0469.2009.00532.x
- 1097 Pinho, Catarina, Ferrand, N., & Harris, D. J. (2006). Reexamination of the Iberian and
1098 North African *Podarcis* (Squamata: Lacertidae) phylogeny based on increased
1099 mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, *38*(1),

- 1100 266–273. doi: 10.1016/j.ympcv.2005.06.012
- 1101 Pinho, Catarina, Harris, D. J., & Ferrand, N. (2007). Contrasting patterns of
1102 population subdivision and historical demography in three western
1103 Mediterranean lizard species inferred from mitochondrial DNA variation.
1104 *Molecular Ecology*, 16(6), 1191–1205. doi: 10.1111/j.1365-294X.2007.03230.x
- 1105 Pinho, Catarina, & Hey, J. (2010). Divergence with gene flow: models and data.
1106 *Annual Review of Ecology, Evolution, and Systematics*, 41, 215–230.
- 1107 Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but
1108 narrows the expected cline width. *Genetics*, 189(1), 227–235.
- 1109 Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., & Chapuisat, M. (2014). Convergent
1110 genetic architecture underlies social organization in ants. *Current Biology*,
1111 24(22), 2728–2732.
- 1112 Puritz, J. B., Matz, M. V, Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E.
1113 (2014). Demystifying the RAD fad. *Molecular Ecology*, 23, 5937–5942.
- 1114 Qvarnström, A., & Bailey, R. I. (2009). Speciation through evolution of sex-linked
1115 genes. *Heredity*, 102(1), 4.
- 1116 Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ...
1117 Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road
1118 map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8),
1119 1450–1477.
- 1120 Rieseberg, L. H. (2009). Evolution: replacing genes and traits through hybridization.
1121 *Current Biology*, 19(3), R119–R122. doi:
1122 <https://doi.org/10.1016/j.cub.2008.12.016>
- 1123 Sá-Sousa, P. (2001). Comparative chorology between *Podarcis bocagei* and *P.*
1124 *carbonellae* (Sauria: Lacertidae) in Portugal. *Revista Española de Herpetología*,

- 1125 15, 85–97.
- 1126 Scascitelli, M., Whitney, K. D., Randell, R. A., King, M., Buerkle, C. A., & Rieseberg,
1127 L. H. (2010). Genome scan of hybridizing sunflowers from Texas (*Helianthus*
1128 *annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small
1129 islands of genomic differentiation. *Molecular Ecology*, 19(3), 521–541.
- 1130 Storchová, R., Reif, J., & Nachman, M. W. (2010). Female heterogamety and
1131 speciation: reduced introgression of the Z chromosome between two species of
1132 nightingales. *Evolution*, 64(2), 456–471. doi: 10.1111/j.1558-5646.2009.00841.x
- 1133 Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between
1134 the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in
1135 southern Poland. *Evolution*, 40(6), 1141–1159.
- 1136 Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone
1137 between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons
1138 between transects and between loci. *Evolution*, 45(2), 237–261.
- 1139 Tarroso, P., Pereira, R. J., Martínez-Freiría, F., Godinho, R., & Brito, J. C. (2014).
1140 Hybridization at an ecotone: Ecological and genetic barriers between three
1141 Iberian vipers. *Molecular Ecology*, 23(5), 1108–1123. doi: 10.1111/mec.12671
- 1142 Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., &
1143 Tucker, P. K. (2010). The variable genomic architecture of isolation between
1144 hybridizing species of house mice. *Evolution: International Journal of Organic*
1145 *Evolution*, 64(2), 472–485.
- 1146 Teeter, K., & Payseur, B. (2008). Genome-wide patterns of gene flow across a house
1147 mouse hybrid zone. *Genome Research*, 18, 67–76. doi: 10.1101/gr.6757907.1
- 1148 Templeton, A. R. (1992). The meaning of species and speciation: a genetic
1149 perspective. *The Units of Evolution*, 159–185.

- 1150 Tucker, P. K., Sage, R. D., Warner, J., Wilson, A. C., & Eicher, E. M. (1992). Abrupt
1151 cline for sex chromosomes in a hybrid zone between two species of mice.
1152 *Evolution*, *46*(4), 1146–1163.
- 1153 Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in
1154 *Anopheles gambiae*. *PLoS Biology*, *3*(9), e285.
- 1155 van Schooten, B., Meléndez-Rosa, J., Van Belleghem, S. M., Jiggins, C. D., Tan, J.
1156 D., McMillan, W. O., & Papa, R. (2020). Divergence of chemosensing during the
1157 early stages of speciation. *Proceedings of the National Academy of Sciences*,
1158 *117*(28), 16438–16447.
- 1159 Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during
1160 ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal*
1161 *Society of London B: Biological Sciences*, *367*(1587), 451–460.
- 1162 Vignoli, L., Vuerich, V., & Bologna, M. A. (2012). Experimental study of dispersal
1163 behaviour in a wall lizard species (*Podarcis sicula*) (Sauria Lacertidae). *Ethology*
1164 *Ecology & Evolution*, *24*(3), 244–256.
- 1165 Walsh, J., Shriver, W. G., Olsen, B. J., & Kovach, A. I. (2016). Differential
1166 introgression and the maintenance of species boundaries in an advanced
1167 generation avian hybrid zone. *BMC Evolutionary Biology*, *16*(1), 1–18.
- 1168 Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary*
1169 *Biology*, *14*(6), 851–865. doi: 10.1046/j.1420-9101.2001.00335.x
- 1170 Yang, W., Feiner, N., Pinho, C., While, G. M., Kaliontzopoulou, A., Harris, D. J., ...
1171 Uller, T. (2021). Extensive introgression and mosaic genomes of Mediterranean
1172 endemic lizards. *Nature Communications*, *12*(1), 1–8.
- 1173

1174 Data Accessibility Statement

1175

1176 Custom Python script used to select one SNP per locus is available
1177 at https://github.com/catpinho/filter_RADseq_data and the scripts used to perform the
1178 simulations, format conversions and data set manipulations are available at
1179 https://github.com/catpinho/admixture_simulation. Raw sequence reads from ddRAD-
1180 seq are deposited in the NCBI Sequence Read Archive (SRA; BioProject
1181 PRJNA665746). BioSample numbers for each individual are provided in
1182 Supplemental Information File 2. Genotype data in VCF format is available on Dryad
1183 (DOI <https://doi.org/10.5061/dryad.n8pk0p31k>). Sequences of mtDNA are deposited
1184 in GeneBank and accession numbers for each individual are provided in
1185 Supplemental Information File 2.

1186

1187 Author Contributions

1188

1189 GCD was responsible for field and laboratory work, bioinformatics, data
1190 analysis, and writing the manuscript. AB was responsible for laboratory
1191 work and manuscript revision. MMR was responsible for field and laboratory
1192 work and manuscript revision. PAC was responsible for project conception,
1193 design and coordination, manuscript revision and obtaining funding. CP was
1194 responsible for project conception, design and coordination, bioinformatics,
1195 data analysis, manuscript writing and revision, and obtaining funding.

1196

1197 Figures

1198

1199 **Figure 1 – a)** Map of Iberian Peninsula with distribution of *Podarcis bocagei* (in red)
 1200 and *P. carbonelli* (in blue); **b)** The highlighted area in a) showing the locations of
 1201 sampled populations in the transect across the hybrid zone; **c)** zoom in of the area in
 1202 b) where individuals of both parental species were found in syntopy. Each triangle
 1203 represents an individual assigned to each parental species ($Q < 0.1$ or $Q > 0.9$) and
 1204 each star represents an individual with Q between 0.1 and 0.9 based on Admixture
 1205 analysis.

1206

1207 **Figure 2 –** Population structure across the hybrid zone between *P. bocagei* and *P.*
 1208 *carbonelli*. **a)** Results from individual multilocus genotype clustering for the complete
 1209 dataset (329 individuals and 6905 loci). Each individual is represented by a vertical
 1210 line proportionally partitioned into the $K=2$ and $K=3$ coloured segments. Population
 1211 acronyms and colours as in Figure 1. **b)** Distribution of Q_c scores in 10 classes of
 1212 equal size within the syntopy area, derived from the multilocus genotype assignment
 1213 analysis with the complete dataset. **c)** Principal Component Analysis (PCA) for the
 1214 complete dataset. Circles represent individuals from populations north of the contact
 1215 zone, triangles correspond to the individuals from the contact zone and squares
 1216 identify the individuals from populations south of the contact zone. Red represents
 1217 individual identified as *P. bocagei* and blue as *P. carbonelli* based on morphological
 1218 identification in the field. **d)** Triangle plot showing the distribution of individual hybrid
 1219 index ($HI_{DIAG-INT}$) and interspecific heterozygosity (H_e) in the contact zones based on
 1220 diagnostic dataset (1241 loci) between reference populations (POR and TOR) in the
 1221 contact zone. Colours and shapes as in panel **b**. *Podarcis bocagei* reference
 1222 individuals were set to have a HI of zero, and *P. carbonelli* was set to one.

1223

1224 **Figure 3 –** Box plots representing the ratios of simulated distributions of F1 hybrids
 1225 by backcrosses (F1/BC) with different migration rates, i.e., the proportions of
 1226 genotypes replaced each generation by parental genotypes ($HI = 0$ or 1) in the same
 1227 proportion as they appear in the observed dataset. **a)** Distributions based on
 1228 simulations with random mating and **b)** based on non-random mating. In both panels
 1229 horizontal dotted lines represent the F1/BC ratio in the observed data.

1230

1231 **Figure 4 –** Geographic and genomic cline variation across SNPs. **a)** Geographic
 1232 cline variation of the 2300 loci (“80/20” dataset) described by c and w , estimated with
 1233 HZAR. Gray circles are the loci without significant differences in c when compared to
 1234 c_{HI} (non-sig.), red circles are the loci with c significantly shifted to North (sign. to
 1235 North) and blue circles are the loci with c significantly shifted to South (sign. to
 1236 South). Yellow filled circles are the loci with w significantly higher than w_{HI} (sign.
 1237 larger). The circles with an asterisk denote loci with outlier c . Horizontal dotted line
 1238 represent the c_{HI} (13.15km) and the vertical represents the w_{HI} (0.54km). Solid line
 1239 represents the regression line between c and w values. Regression significance is
 1240 displayed on top-right corner. **b)** Genomic cline variation in the contact zone (ESP
 1241 population) described by α and β , estimated with BGC. Each point represents one of
 1242 the 6905 loci (complete dataset), and contour lines show regions of
 1243 the α - β parameter space with similar density of loci.

1244

1245 **Figure 5** – Concordance of analogous parameters from geographic and genomic
 1246 clines using the 2300 loci common to both complete and “80/20” datasets.
 1247 Regression lines, it’s respective r^2 values and their significance are shown in each
 1248 plot. **a)** Concordance of α and c . Circles filled in red are loci with α significantly
 1249 associated *P. bocagei* genomic background and circles filled in blue are loci with α
 1250 significantly associated *P. carbonelli* genomic background. Grey filled circles are loci
 1251 that α do not significantly differ from neutral expectations. Red circles represent loci
 1252 with c significantly shifted from c_{HI} to North and blue circles to South. Grey circles are
 1253 loci that c do not significantly differ from c_{HI} . **b)** Concordance of β and w . Yellow filled
 1254 circles represent loci with increased introgression rate (significant negative values)
 1255 and green with decreased introgression rate (significant positive values), relatively to
 1256 neutral expectations. Grey filled circles are loci that β do not significantly differ from
 1257 neutral expectations. Yellow circles are loci with w significantly higher from w_{HI} (loci
 1258 with w significantly lower from w_{HI} were not detected) and grey circles are loci that w
 1259 do not significantly differ from w_{HI} .

1260
 1261 **Figure 6** – Distribution of geographic and genomic cline parameters across the loci
 1262 mapped on the *P. muralis* genome. **a)** Genomic distribution of loci and it’s respective
 1263 geographic and genomic cline parameters estimated with HZAR and BGC,
 1264 respectively. Each grey bar in the bottom represents one chromosome and it is
 1265 proportional to its size. Each coloured dot represents statistically significant loci for
 1266 each parameter. Colours as in Figure 5. **b, c)** Violin plots representing the density of
 1267 values for geographic cline centre (c) and width (w) estimated with HZAR for the
 1268 2101 mapped loci contained in the “80/20” dataset. Due to extreme outlier values in
 1269 both panels, y-axis is not continuous to facilitate visualization; plots with complete set
 1270 of outliers are shown in Supplemental Information File 1, Figure S1. **d, e)** Violin plots
 1271 representing the density of values for genomic cline centre (α) and introgression rate
 1272 (β) estimated with BGC for the 6027 mapped loci contained in the complete dataset.
 1273 The horizontal solid line within each box plot represents the median.

1274 **Tables**

1275
 1276
 1277 **Table 1** – Summary results of geographic cline analysis performed with HZAR using
 1278 2300 loci (“80/20” dataset, upper section) and genomic cline analysis performed with
 1279 BGC using 6905 loci (complete dataset, lower section). For each section are
 1280 represented the number of loci (N) and correspondent percentage falling in each class
 1281 of geographic clines centre (c)/cline width (w) and locus specific ancestry
 1282 (α)/introgression rate (β) of genomic cline.

1283
 1284 **Table 2** – Non-parametric global Kruskal-Wallis tests to assess if the distribution of
 1285 each parameter value used to describe introgression, from both geographic (c and w)
 1286 or genomic (α and β) cline analyses, were similar between chromosomes. Significant
 1287 differences ($p < 0.05$) are marked with an asterisk (*).

1288

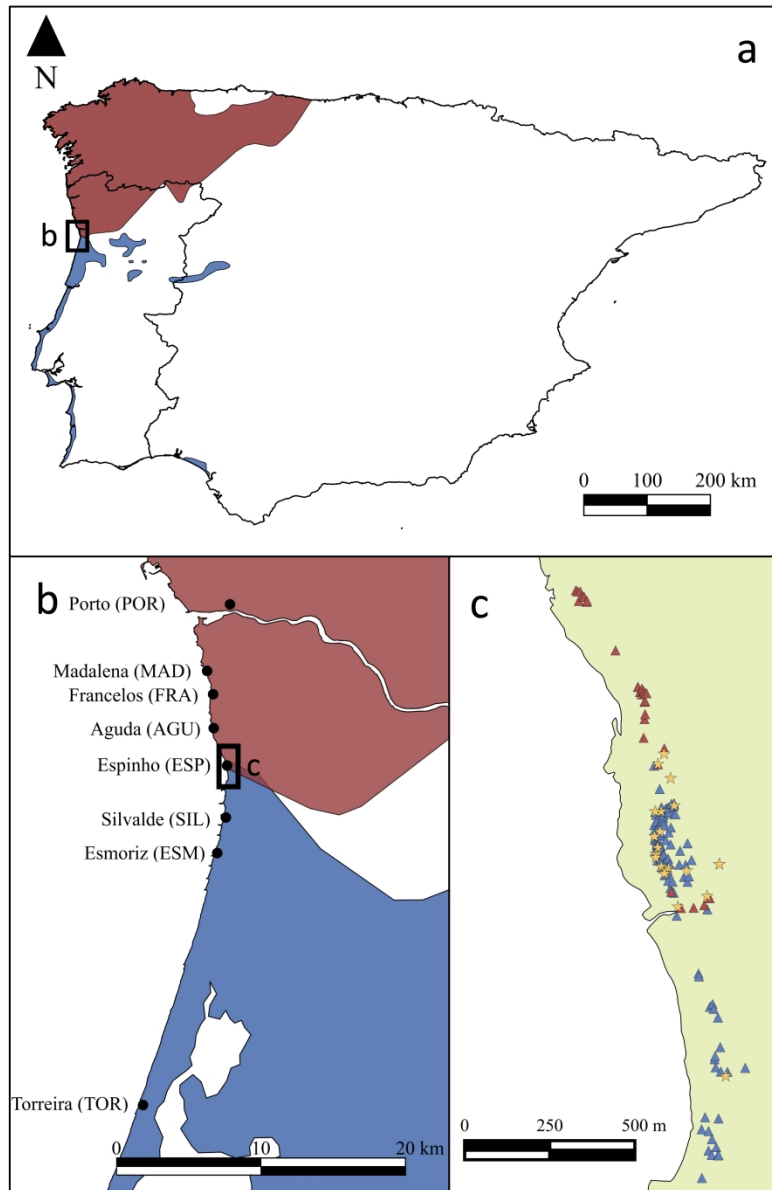


Figure 1 – a) Map of Iberian Peninsula with distribution of *Podarcis bocagei* (in red) and *P. carbonelli* (in blue); b) The highlighted area in a) showing the locations of sampled populations in the transect across the hybrid zone; c) zoom in of the area in b) where individuals of both parental species were found in syntopy. Each triangle represents an individual assigned to each parental species ($Q < 0.1$ or $Q > 0.9$) and each star represents an individual with Q between 0.1 and 0.9 based on Admixture analysis.

1276x1937mm (72 x 72 DPI)

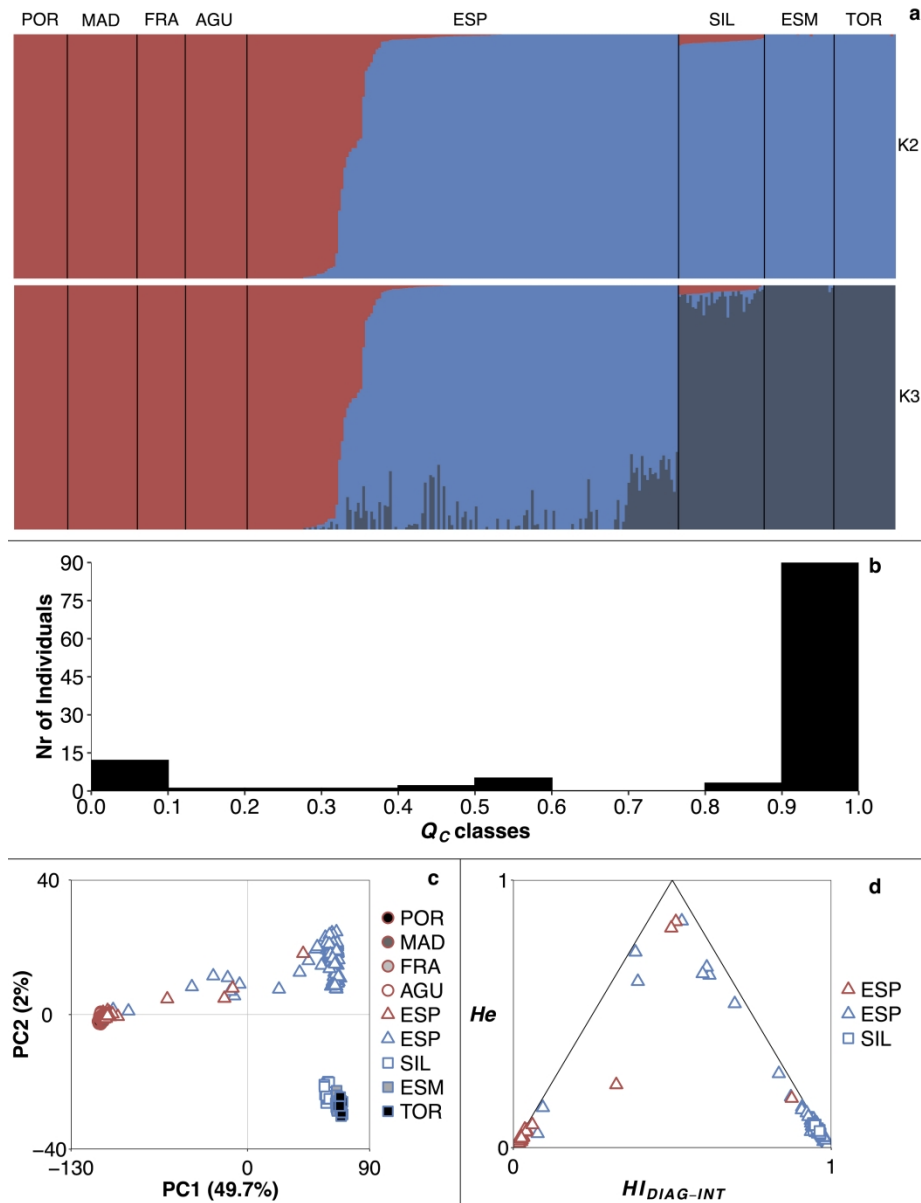


Figure 2 – Population structure across the hybrid zone between *P. bocagei* and *P. carbonelli*. a) Results from individual multilocus genotype clustering for the complete dataset (329 individuals and 6905 loci). Each individual is represented by a vertical line proportionally partitioned into the K=2 and K=3 coloured segments. Population acronyms and colours as in Figure 1. b) Distribution of Q_c scores in 10 classes of equal size within the syntopy area, derived from the multilocus genotype assignment analysis with the complete dataset. c) Principal Component Analysis (PCA) for the complete dataset. Circles represent individuals from populations north of the contact zone, triangles correspond to the individuals from the contact zone and squares identify the individuals from populations south of the contact zone. Red represents individual identified as *P. bocagei* and blue as *P. carbonelli* based on morphological identification in the field. d) Triangle plot showing the distribution of individual hybrid index ($HI_{DIAG-INT}$) and interspecific heterozygosity (He) in the contact zones based on diagnostic dataset (1241 loci) between reference populations (POR and TOR) in the contact zone. Colours and shapes as in panel b. *Podarcis bocagei* reference individuals were set to have a HI of zero, and *P. carbonelli* was set to one.

2114x2774mm (65 x 65 DPI)

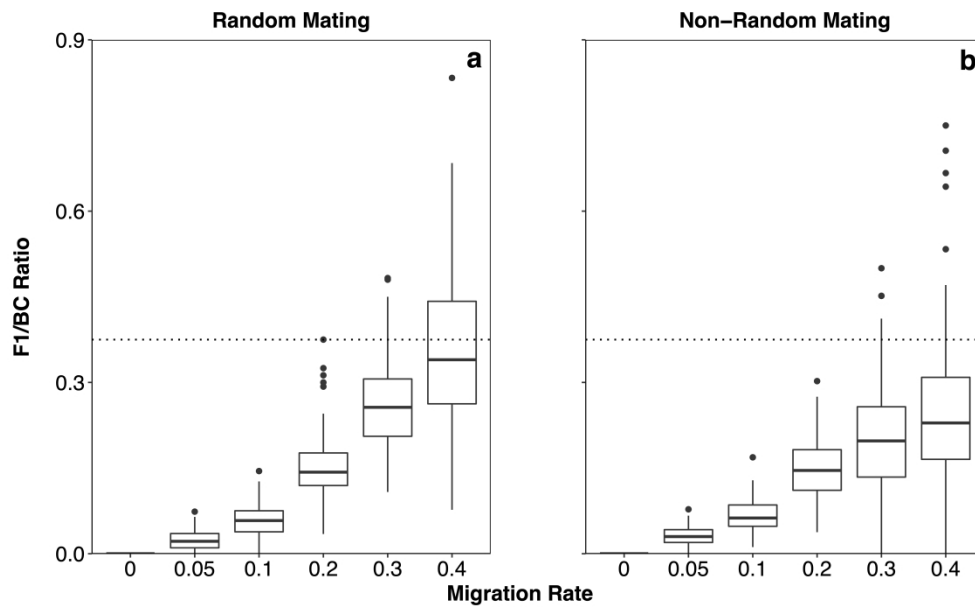


Figure 3 – Box plots representing the ratios of simulated distributions of F1 hybrids by backcrosses (F1/BC) with different migration rates, i.e., the proportions of genotypes replaced each generation by parental genotypes ($HI = 0$ or 1) in the same proportion as they appear in the observed dataset. a) Distributions based on simulations with random mating and b) based on non-random mating. In both panels horizontal dotted lines represent the F1/BC ratio in the observed data.

2363x1439mm (72 x 72 DPI)

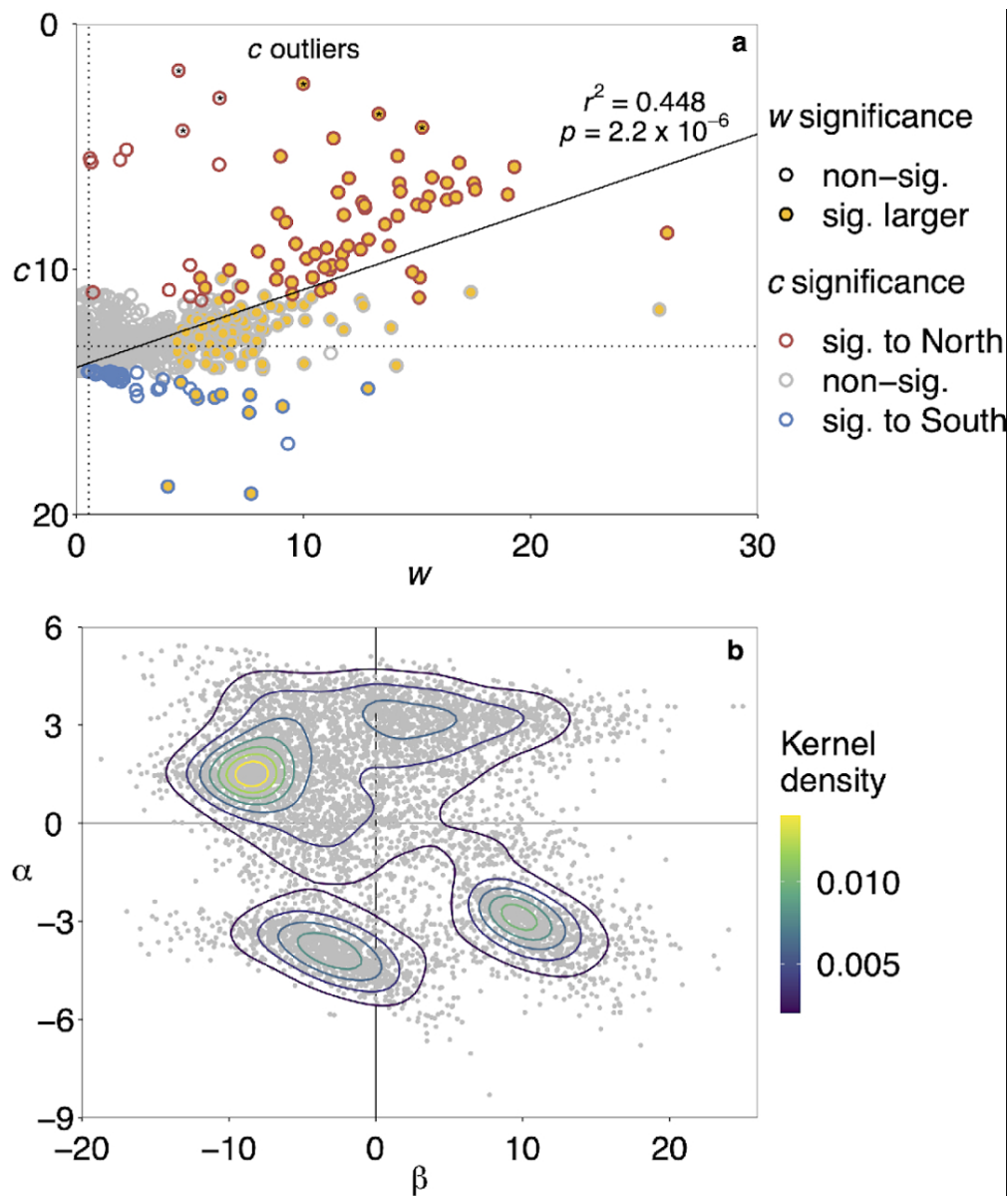


Figure 4 – Geographic and genomic cline variation across SNPs. a) Geographic cline variation of the 2300 loci ("80/20" dataset) described by c and w , estimated with HZAR. Gray circles are the loci without significant differences in c when compared to c_{HI} (non-sig.), red circles are the loci with c significantly shifted to North (sign. to North) and blue circles are the loci with c significantly shifted to South (sign. to South). Yellow filled circles are the loci with w significantly higher than w_{HI} (sign. larger). The circles with an asterisk denote loci with outlier c . Horizontal dotted line represent the c_{HI} (13.15km) and the vertical represents the w_{HI} (0.54km). Solid line represents the regression line between c and w values. Regression significance is displayed on top-right corner. b) Genomic cline variation in the contact zone (ESP population) described by α and β , estimated with BGC. Each point represents one of the 6905 loci (complete dataset), and contour lines show regions of the α – β parameter space with similar density of loci.

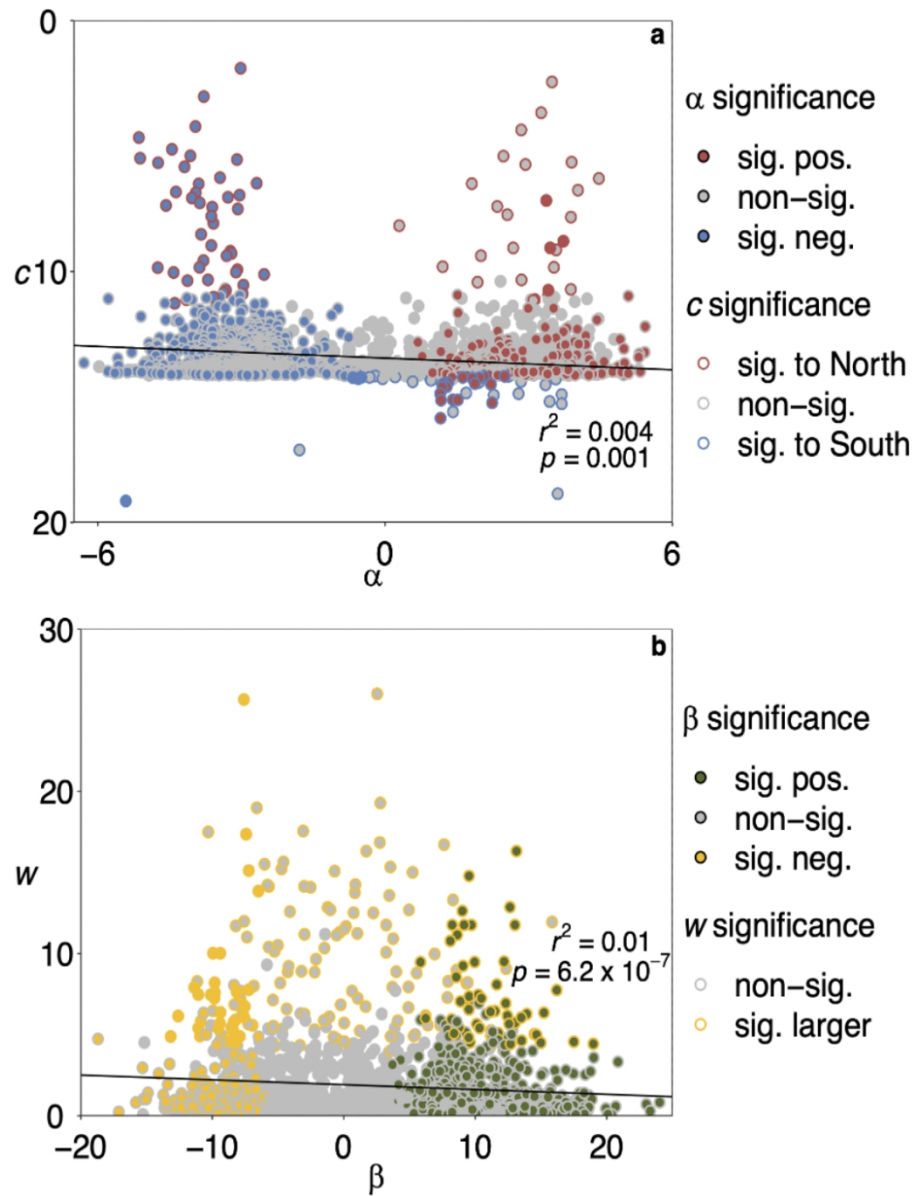


Figure 5 – Concordance of analogous parameters from geographic and genomic clines using the 2300 loci common to both complete and “80/20” datasets. Regression lines, it’s respective r^2 values and their significance are shown in each plot. a) Concordance of a and c . Circles filled in red are loci with a significantly associated *P. bocagei* genomic background and circles filled in blue are loci with a significantly associated *P. carbonelli* genomic background. Grey filled circles are loci that a do not significantly differ from neutral expectations. Red circles represent loci with c significantly shifted from c_{HI} to North and blue circles to South. Grey circles are loci that c do not significantly differ from c_{HI} . b) Concordance of β and w . Yellow filled circles represent loci with increased introgression rate (significant negative values) and green filled circles represent loci with decreased introgression rate (significant positive values), relatively to neutral expectations. Grey filled circles are loci that β do not significantly differ from neutral expectations. Yellow circles are loci with w significantly higher from w_{HI} (loci with w significantly lower from w_{HI} were not detected) and grey circles are loci that w do not significantly differ from w_{HI} .

188x248mm (144 x 144 DPI)

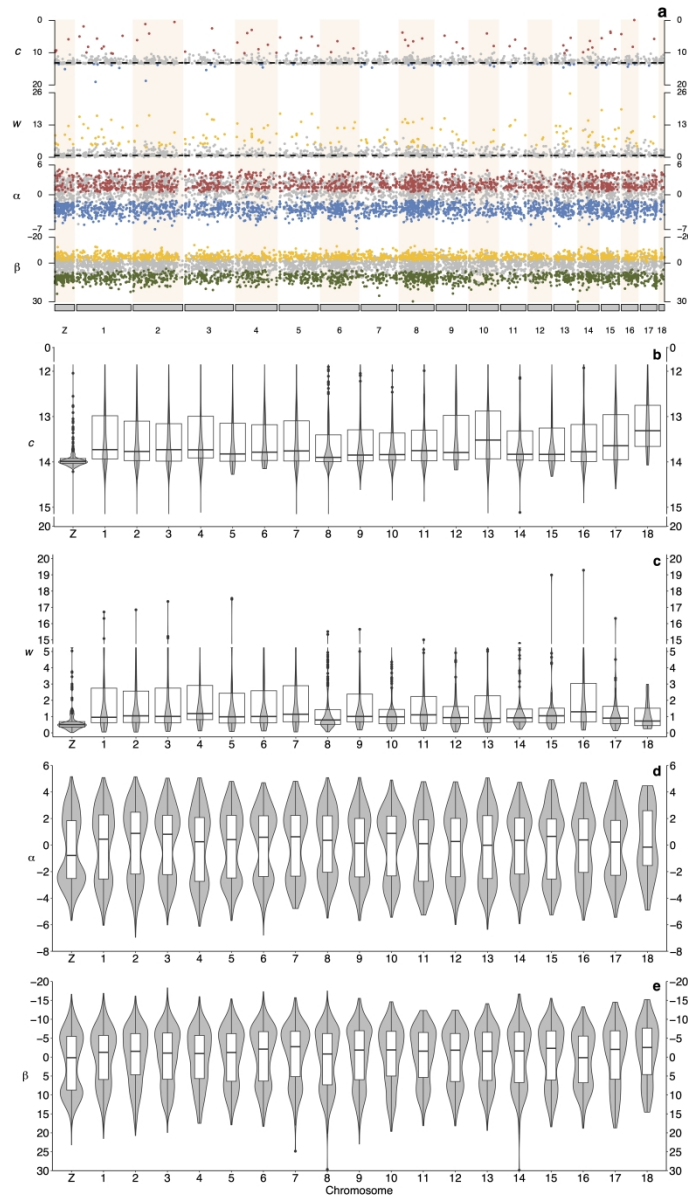


Figure 6 – Distribution of geographic and genomic cline parameters across the loci mapped on the *P. muralis* genome. a) Genomic distribution of loci and its respective geographic and genomic cline parameters estimated with HZAR and BGC, respectively. Each grey bar in the bottom represents one chromosome and it is proportional to its size. Each coloured dot represents statistically significant loci for each parameter. Colours as in Figure 5. b, c) Violin plots representing the density of values for geographic cline centre (c) and width (w) estimated with HZAR for the 2101 mapped loci contained in the “80/20” dataset. Due to extreme outlier values in both panels, y-axis is not continuous to facilitate visualization; plots with complete set of outliers are shown in Supplemental Information File 1, Figure S1. d, e) Violin plots representing the density of values for genomic cline centre (α) and introgression rate (β) estimated with BGC for the 6027 mapped loci contained in the complete dataset. The horizontal solid line within each box plot represents the median.

1625x2810mm (72 x 72 DPI)

Geographic clines		w							
		Similar to <i>HI</i>		Smaller than <i>HI</i>		Higher than <i>HI</i>		Total	
		N	%	N	%	N	%	N	%
c	Not shifted	2064	89.7	0	0	116	5	2180	94.8
	Shifted North	13	0.6	0	0	60	2.6	73	3.2
	Shifted South	36	1.6	0	0	11	0.5	47	2
	Total	2113	91.9	0	0	187	8.1	2300	100

Genomic clines		β							
		Zero		Negative		Positive		Total	
		N	%	N	%	N	%	N	%
α	Zero	3194	46.3	354	5.1	77	1.1	3625	52.5
	Positive	205	2.9	755	10.9	93	1.3	1053	15.2
	Negative	1068	15.5	37	0.5	1122	16.2	2227	32.3
	Total	4467	64.7	1146	16.7	1292	18.7	6905	100

Parameter		Kruskal-Wallis chi-squared value	<i>p</i> -value
Geographic clines	<i>c</i>	169.4568103	1.16x10 ⁻²⁸ *
	<i>w</i>	186.3242633	5.33x10 ⁻³⁰ *
Genomic clines	<i>α</i>	22.10320156	0.009 *
	<i>β</i>	33.67921046	8.43x10 ⁻⁰⁸ *