

Can secondary contact following range expansion be distinguished from barriers to gene flow?

Johanna Bertl¹ Michael G. B. Blum^{2,3}

March 12, 2016

¹ Department of Molecular Medicine (MOMA), Aarhus University, Denmark

² Université Grenoble Alpes, Laboratoire TIMC-IMAG, UMR 5525, Grenoble, France.

³ CNRS, TIMC-IMAG, Grenoble, France.

Abstract

Secondary contact is the reestablishment of gene flow between sister populations that have diverged. For instance, at the end of the Quaternary glaciations in Europe, secondary contact occurred during the northward expansion of the populations which had found refugia in the southern peninsulas. With the advent of multi-locus markers, secondary contact can be investigated using various molecular signatures including gradients of allele frequency, admixture clines, and local increase of genetic differentiation. We use coalescent simulations to investigate if molecular data provide enough information to distinguish between secondary contact following range expansion and an alternative evolutionary scenario consisting of a barrier to gene flow in an isolation-by-distance model. Although evidence for secondary contact is usually conveyed by statistics related to admixture coefficients, we find that they have no power to make the distinction. By contrast, the directionality index ψ that was proposed to study range expansion is informative. Additionally, we find that an excess of Linkage Disequilibrium and of genetic diversity at the suture zone is a unique signature of secondary contact. Our findings indicate that inference on secondary contact can be improved when explicitly accounting for the geographical locations of individuals.

1 Introduction

2 Hybrid zones are narrow regions in which genetically distinct populations meet,
3 mate, and produce hybrids (Barton and Hewitt, 1985). Hybrid zones induced
4 by secondary contact have often been observed in connection with the Qua-
5 ternary glaciations (Hewitt, 2000). For instance, molecular markers suggest
6 that the southern peninsulas of Europe were major ice age refugia of the Euro-
7 pean biota and that secondary contact occurred during the northward expansion
8 which followed the last glacial maximum (Taberlet et al., 1998; Hewitt, 1999).
9 With the advent of multi-locus molecular markers such as microsatellite or SNP
10 data, hybrid zones can be investigated using various molecular signatures in-
11 cluding gradients of allele frequency, admixture clines, and local increase of
12 genetic differentiation (Nielsen et al., 2003; Adams et al., 2006; Strand et al.,
13 2012; Bermond et al., 2012). Molecular or morphological clinal patterns pro-
14 vide evidence for secondary contact in various plant and animal species such
15 as *Arabidopsis thaliana* (Huber et al., 2014), *Silene vulgaris* (Keller and Taylor,
16 2010), the grasshopper *Oedaleus decorus* (Kindler et al., 2012), the European
17 hare *Lepus europaeus* or the parrotbill bird *Paradoxornis webbianus* (Qu et al.,
18 2012) to name just a few examples.

19 However, typical molecular signatures of secondary contact zones can also
20 occur under other evolutionary scenarios. For instance, admixture clines can be
21 observed under pure isolation-by-distance models where nearby populations are
22 connected through gene flow (Engelhardt and Stephens, 2010). Additionally,
23 an increase of genetic differentiation can occur in isolation-by-distance models
24 when there are barriers to dispersal (Riley et al., 2006). With the advent of land-
25 scape genetics, the search for barriers to gene flow has attracted considerable
26 attention (Manel et al., 2003; Storfer et al., 2010). Although secondary-contact
27 zones and barriers to gene flow are not mutually exclusive, because secondary

28 contact is a non-equilibrium situation that converges to a migration-drift equi-
29 librium (Endler, 1977), they convey different evolutionary paradigms. Models of
30 barriers to gene flow are usually based on isolation-by-distance settings where
31 neighboring populations are connected through dispersal (Safner et al., 2011;
32 Blair et al., 2012). Around the barrier to gene flow, dispersal is lowered be-
33 cause of geographical or anthropogenic obstacles (Riley et al., 2006; Zalewski
34 et al., 2009). By contrast, models of secondary-contact include an initial phase
35 of evolutionary divergence between two populations or between two sets of pop-
36 ulations. The phase of evolutionary divergence is followed by a phase of gene
37 flow between the two divergent units at the secondary contact zone (Murray
38 and Hare, 2006; Durand et al., 2009). Here, we use coalescent simulations to
39 investigate to what extent molecular data provide information to distinguish
40 between the two alternatives. For both evolutionary scenarios, we consider a
41 one-dimensional range with isolation-by-distance as shown in figures 1 and 2.

42 For comparing the molecular signal left by the two distinct scenarios, we
43 consider statistical measures that have been developed to provide evidence for
44 different demographic processes. The first set of summary statistics we ex-
45 plore is typically used to **detect** hybrid zones. This first set contains measures
46 of individual admixture coefficients between the parental source populations
47 (Nielsen et al., 2003; Durand et al., 2009), a statistical test based on Linkage
48 Disequilibrium (LD) where we explicitly test if a population results from ad-
49 mixture between parental source populations (Chakraborty and Weiss, 1988),
50 and a measure of LD as we expect an increase of LD in admixed populations
51 (McVean, 2002).

52 Secondary contact is frequently induced by geographical expansions of the
53 ancestral populations (Hewitt, 2000), so the second set of summary statistics
54 corresponds to measures of evidence for range expansion. We consider the re-

55 cently developed directionality index ψ as it is sensitive to the occurrence of
56 recent range expansion and it should distinguish between equilibrium and non-
57 equilibrium models (Peter and Slatkin, 2013). The properties of the direc-
58 tionality index have not been studied yet when there are introgressive events.
59 Furthermore, we include genetic diversity which has been shown to decrease
60 along the expansion direction (Austerlitz et al., 1997).

61 The last set of summary statistics pertains to isolation-by-distance and barriers
62 to gene flow. Numerical summaries indicative of barriers to gene flow usually
63 measure genetic discontinuities, which are zones of sharp changes in allele fre-
64 quencies (Manel et al., 2003). There are different ways to detect and measure
65 them. Here, we use local F_{ST} defined as F_{ST} per unit of spatial distance. Local
66 F_{ST} can be provided from multiple loci with georeferenced data by the soft-
67 ware *LocalDiff* and we expect them to be larger around the barrier to gene flow
68 (Duforet-Frebourg and Blum, 2014). The development of the software *LocalDiff*
69 questioned the possibility of distinguishing between the two evolutionary scenar-
70 ios under consideration (Duforet-Frebourg and Blum, 2014). Last, we include
71 the decay of correlation between allele frequencies as a function of distance as
72 it provides evidence of isolation-by-distance (Hardy and Vekemans, 1999).

73 2 Methods

74 2.1 Models

75 We consider secondary contact in a one-dimensional nearest-neighbor stepping-
76 stone model consisting of 100 demes (figure 1). Range expansion is modeled as a
77 series of founder events with moderate bottlenecks. Time is given in coalescent
78 units before present time, i. e. in units of $2N$ generations where N is the diploid
79 population size per deme at present time. Accordingly, all parameters are scaled

80 with $4N$.

81 **Phase 1 (ancestral population)** The ancestral population is a random-mating
82 population of size $2N$. At time t_S , it splits in two populations of size N .

83 **Phase 2 (separate refugia)** From time t_S to time t_E , the two populations
84 are in separate refugia (demes 1 and 100, respectively), the population
85 sizes are constant, and there is no gene flow.

86 **Phase 3 (expansion)** Starting at time t_E , both populations expand towards
87 each other in the stepping-stone geometry. At time points t_E, t_E-d, t_E-2d
88 etc., 10% of the individuals of the deme at the expansion front colonize a
89 new deme. Instantaneously, the size of both demes increases to N again
90 and migration occurs at rate μ between neighboring demes.

91 **Phase 4 (secondary contact)** From $t_C = t_E - 48d$ until the present time,
92 a stepping-stone model with 100 demes of size N is maintained with mi-
93 gration rate μ among the neighboring demes including demes 50 and 51
94 where secondary contact occurs.

95 As an alternative we investigate a nearest-neighbor stepping-stone model
96 with a constant range of 100 demes and with reduced gene flow in the center
97 (figure 2). The barrier to gene flow is modeled by a lower migration rate $\mu_B \leq \mu$
98 between demes 50 and 51.

99 DNA data of 20 chromosomes per deme is simulated with the coalescent
100 simulator `ms` (Hudson, 2002). On each chromosome, we simulate 100 unlinked
101 sequences consisting of 100,000 base pairs each. A sequence contains 100 SNPs
102 and the scaled recombination rate within the sequence is 4.

103 In the secondary contact model we simulate data with parameters $t_S = 19$,
104 $d = 1/8$ and different durations since secondary contact occurred (from $t_C =$
105 $0, 1, \dots, 5$ until present time). In both models the scaled migration rate between

106 neighboring demes is $\mu = 20$. In the stepping-stone model we consider different
107 barrier permeabilities ($\mu_B/\mu = 0.002, 0.01, 0.02, 0.1, 0.2, 1$; a value of 1 denotes
108 no barrier). To provide means and standard errors of the summary statistics,
109 each simulation is repeated 100 times.

110 2.2 Summary statistics

111 Hybrid zone summary statistics

112 **Admixture coefficient** Based on the first principal component we compute
113 an admixture coefficient for the pool of the five demes left of the barrier (Paschou
114 et al., 2007; Bryc et al., 2010). The pools of the 5 leftmost and 5 rightmost demes
115 are used as proxies for the two source populations. The admixture coefficient is
116 defined as the average relative location of individuals in the putative admixed
117 population with respect to the two source populations on the axis of the first
118 principal component. It takes values between 0 and 1 and the larger it is,
119 the higher is the proportion of genetic material inherited from the right source
120 population (across the barrier or hybrid zone) through admixture.

121 The principal component analysis is conducted with the R function `prcomp`
122 (R Core Team, 2012).

A test for admixture based on Linkage Disequilibrium (LD) Let δ_i
be the allele frequency difference between the two source populations at locus i .
LD in the admixed population after t generations of admixture with migration
rate m is given by

$$LD^{(t)} = (1 - \rho)^t m(1 - m)\delta_1\delta_2,$$

123 where ρ is the recombination rate between loci 1 and 2 (Chakraborty and Weiss,
124 1988, eq. 3).

125 A significant correlation between LD and the product of differences of allele

126 frequencies can be taken as a statistical evidence for admixture (Bray et al.,
127 2010). Thus, we test admixture by computing P-values based on the Spearman
128 correlation between LD and $\delta_1\delta_2$ using all pairs of SNPs that are part of the same
129 sequence. The admixed and source populations are defined as for computing
130 the admixture coefficient.

131 **Linkage disequilibrium (LD)** The squared correlation coefficient between
132 1,000 randomly drawn pairs of SNPs within the same sequence is averaged over
133 all unlinked sequences. LD is computed for each deme.

134 **Range expansion summary statistics**

135 **Directionality index ψ** The directionality index ψ has been defined to detect
136 a range expansion and infer its origin (Peter and Slatkin, 2013). The index
137 uses the fact that populations further away from the origin of expansion have
138 experienced more genetic drift. The index $\psi_{i,j}$ is a pairwise measure between
139 demes i and j that compares the average allele frequencies in the two demes:
140 stronger drift yields higher differences in allele frequencies. Given that a range
141 expansion has occurred, $\psi_{i,j}$ should be negative if i is closer to the origin of the
142 expansion than j , and positive otherwise. If $\psi_{i,j} \approx 0$, both demes should be
143 equally close to the origin of the expansion, or no range expansion has occurred.
144 We compute values of $\psi_{26,j}$ with $j = 27, \dots, 50$.

145 **Genetic variability** The genetic variability is measured in each deme by the
146 average number of pairwise nucleotide differences between all pairs of sequences,
147 denoted by Δ .

148 **Summary statistics for isolation-by-distance and barriers to gene flow**

149 **Allele frequency correlogram** The Pearson correlation between the allele
150 frequencies of demes i and j is denoted by $r_{i,j}$. We compute the correlogram
151 $r_{26,j}$ for $j = 27, \dots, 75$.

152 **Local F_{ST}** Local values of F_{ST} correspond to pairwise F_{ST} between neigh-
153 boring demes (Duforet-Frebourg and Blum, 2014).

154 **3 Results**

155 For the two different scenarios, we plot the summary statistics either as a func-
156 tion of the time since secondary or as a function of the intensity of gene flow
157 across the barrier (figure 3). For the summary statistics that are computed per
158 deme (genetic diversity Δ , LD) or per pair of neighboring demes (F_{ST}), we con-
159 sider the pattern along the whole range of demes. The most important features
160 are captured by the ratio between the values at the barrier or the suture zone,
161 respectively, and the values to the left and right of it (see supplementary figures
162 S1 and S2 for examples of the pattern along the whole range of demes).

163 First, we consider the average admixture coefficient for the 5 populations
164 that are located on the left-hand side of the barrier (demes 45–50). For the
165 isolation-by-distance model with a barrier, these 5 populations are found to be
166 admixed to an extent depending on the barrier permeability: when increasing
167 the barrier permeability, admixture coefficients of individuals on the left-hand
168 side of the barrier approach 50%. As expected, the populations are also found to
169 be admixed in the secondary contact model (between 35% and 50%) except for
170 the scenario where data is collected just before secondary contact occurs ($t_C =$
171 0). When we consider the test for admixture based on Linkage Disequilibrium
172 (table 1), we find that it has a power of 100% ($\alpha = 5\%$) to detect admixture after

173 secondary contact ($t_C > 0$). However, the test also rejects the null hypothesis
174 of no admixture in the isolation-by-distance model very often (power $\geq 80\%$ for
175 $\mu_B/\mu \geq 0.01$). Even for the weakest barrier permeability ($\mu_B/\mu = 0.002$), the
176 null hypothesis is rejected in 62 out of 100 simulations.

177 The ratio between LD at the center (demes 49–52) and on both sides of
178 the range (demes 24–27 and demes 74–77; demes closer to the edge of the
179 range are skipped to avoid the edge effect; see supplementary figures S1 and
180 S2) shows that LD is homogeneous along the whole range of demes for different
181 barrier permeabilities in the stepping stone model with a slight increase at the
182 barrier for low values of μ_B/μ . However, in the secondary contact model, LD is
183 considerably increased in the secondary contact zone ranging from a more than
184 2 fold to an approximately 1.3 fold increase. The excess of LD in the center
185 decreases as time since secondary contact increases.

186 Apart from random fluctuations, the directionality index ψ is constant for
187 the stepping-stone model with constant migration rate ($\mu_B/\mu = 1$) as well as for
188 very old secondary contact ($t_C = 5$). More recent secondary contact results in
189 a U-shaped pattern. The pairwise statistics $\psi_{26,i}$, $i = 27, \dots, 50$ first decreases
190 as expected when moving away from the origin of the expansion but it increases
191 again towards the location of secondary contact. For the barrier model with
192 a moderate and strong barrier ($\mu_B/\mu \geq 0.1$), $\psi_{26,i}$ is constant for most of the
193 range, but decreases slightly close to the barrier.

194 Under the stepping stone model, the number of pairwise differences (Δ) is
195 approximately constant over the range of demes and hardly affected by the
196 barrier. Conversely, in the secondary contact model Δ increases in the suture
197 zone, but with more time elapsed since secondary contact it evens out along the
198 whole range. Only when secondary contact has not occurred yet ($t_C = 0$), the
199 statistic Δ captures the effect of range expansion, which decreases the genetic

200 diversity when moving away from the origin.

201 When considering the decay of allele frequency correlation as a function of
202 distance, we find a sharp decrease around the suture zone or around the barrier,
203 respectively. In the stepping-stone model, the correlation decreases linearly with
204 distance and it drops sharply at the barrier, whereas in the secondary contact
205 model we observe a sigmoid shape. For older secondary contact, the sigmoid
206 decay converges towards the linear decay of the pure stepping-stone model.

207 Pairwise F_{ST} between neighboring demes is increased at the barrier to gene
208 flow. The less permeable is the barrier, the larger is F_{ST} at the barrier compared
209 to the rest of the range. In the secondary contact model, local F_{ST} is increased
210 at the center when measured just before secondary contact occurs ($t_C = 0$),
211 but it is constant along the range of demes when secondary contact is already
212 established ($t_C \geq 1$).

213 4 Discussion

214 Our results show that statistics related to admixture coefficients do not provide
215 sufficient evidence for secondary contact following the isolation of populations
216 in the presence of isolation-by-distance. The fact that isolation-by-distance seri-
217 ously affects the ascertainment of population structure has already been largely
218 documented (Novembre and Stephens, 2008; Frantz et al., 2009). Additional
219 summaries of the data such as local values of F_{ST} or decay of correlation with
220 distance were not more informative to support the occurrence of secondary con-
221 tact.

222 By contrast, both an excess of Linkage Disequilibrium and genetic diversity
223 at the suture zone are found to be unique signatures of secondary contact and
224 informative about the timing of the establishment of secondary contact. Al-
225 though the peak of diversity was initially thought to occur in glacial refugia

226 (Hewitt, 2000), it has already been observed that the genetically most diverse
227 populations were not located in southern Europe but at intermediate latitudes
228 resulting from the admixture of divergent lineages that had expanded from sep-
229 arate refugia (Petit et al., 2003). Last, the directionality index ψ that has been
230 proposed to detect range expansion has a distinctive U-shaped pattern when
231 secondary contact follows range expansions (Peter and Slatkin, 2013). This
232 statistic adds to the toolbox of population geneticists and provides one of the
233 first attempts to distinguish between equilibrium and non-equilibrium spatial
234 processes.

235 In our simulations we observe the re-establishment of the equilibrium state
236 when secondary contact is old enough (Barton and Hewitt, 1985): eventually,
237 all traces of secondary contact are lost and all summary statistics converge to
238 the pattern of the stepping stone model (figure 3).

239 If secondary contact occurs at a barrier to gene flow, the difficulty of de-
240 tecting the secondary contact from molecular data is increased. We consider
241 additional simulations where secondary contact occurs in a region where gene
242 flow is reduced by a factor of 10 ($\mu_B/\mu = 0.1$). In many respects, we see an inter-
243 mediate pattern between the two previously considered scenarios, yet, genetic
244 diversity, Linkage Disequilibrium and the directionality index ψ still provide
245 evidence for secondary contact (supplementary figure S6).

246 Our simulation setting was designed to mimic the evolutionary history of
247 species that have undergone a population split during the Quaternary glaciations
248 with subsequent expansion and secondary contact. Assuming a generation time
249 of 1 year and 1,000 diploid organisms per deme, it includes the time frame of
250 expansion and secondary contact after the last glacial maximum in Europe.
251 Species that had spent the last glacial period in southern refugia started to
252 expand northwards around 16,000 years ago, and subsequently, many plants

253 established a stable distribution around 6,000 years ago (Hewitt, 1999). We
254 assume the ancestral population split up and started diverging 38,000 years ago
255 ($t_S = 19$) and the onset of the expansion varies from 16,000 to 6,000 years
256 ago and lasted 6,000 years. Finally, secondary contact is established on the
257 range of 10,000 years ago ($t_C = 5$) to present time ($t_C = 0$; in this setting,
258 both populations have expanded, but no gene-flow has occurred yet). Our
259 simulations show that the molecular signal of secondary contact vanishes after
260 approximately 10,000 years.

261 To apply our results to a particular organism, parameters like the effective
262 population size, the time of divergence and the expansion rate need to be cali-
263 brated. To assess the robustness of our results, we performed simulations of less
264 extreme scenarios. We found that the same summary statistics are distinctive
265 even with more moderate founder events, a lower expansion speed and higher
266 migration rate between demes (supplementary section 2). However, we also ob-
267 serve that the footprint of secondary contact is more difficult to detect when the
268 expansion is very slow or migration rates between neighboring demes are very
269 high (supplementary figures S4 and S5). Even with these parameter settings
270 where secondary contact is more difficult to detect, the directionality index ψ is
271 a discriminant statistic and can detect the direction of a very slow expansion.

272 Our findings are relevant when investigating modes of speciation using com-
273 putational approaches (Becquet and Przeworski, 2009). Secondary contact fol-
274 lowing divergence without gene flow (allopatry) is often compared to models of
275 speciation where species start diverging while exchanging migrants (sympatry
276 or parapatry) (Becquet and Przeworski, 2009; Duvaux et al., 2011; Roux et al.,
277 2013). The different frameworks to study speciation are based on isolation-
278 and-migration models, which do not account for the spatial and potentially
279 continuous repartition of individuals (Pinho and Hey, 2010). As shown in the

280 simulation study, accounting for spatial processes provides additional informa-
281 tion that can partly be caught with the ψ directionality index, which has power
282 to reveal evolutionary events such as secondary contact and range expansions.
283 Accounting for space is a general recommendation that also stands when study-
284 ing admixture between divergent populations of the same species (Patterson
285 et al., 2012). Although isolation-by-distance is usually perceived as a confound-
286 ing factor (Meirmans, 2012), the spatial sampling of individuals is in fact a
287 chance to develop more powerful statistical approaches in evolutionary biology.
288 Accounting for continuous populations should also be possible when performing
289 simulations to choose the most probable scenario of speciation (Duvaux et al.,
290 2011). Numerical simulators of genetic variation that account for the spatial
291 repartitions of individuals are now available (Ray et al., 2010; Kelleher et al.,
292 2013) and should encourage to study speciation models that reflect the complex
293 spatio-temporal dynamics of species evolutionary histories (Alvarado-Serrano
294 and Hickerson, 2015).

295 **Acknowledgments**

296 J. B. was supported by the Vienna Graduate School of Population Genetics
297 (Austrian Science Fund (FWF): W1225-B20) and worked on this project while
298 employed at the Department of Statistics and Operations Research, University
299 of Vienna, Austria.

300 The computational results presented have been achieved using the Vienna
301 Scientific Cluster (VSC) and the GenomeDK HPC cluster at Aarhus University.

302 References

- 303 Adams, S. M., J. B. Lindmeier, and D. D. Duvernell, 2006. Microsatellite
304 analysis of the phylogeography, Pleistocene history and secondary contact
305 hypotheses for the killifish, *Fundulus heteroclitus*. *Mol. Ecol.* 15:1109–1123.
- 306 Alvarado-Serrano, D. F. and M. J. Hickerson, 2015. Spatially explicit summary
307 statistics for historical population genetic inference. *Methods Ecol. Evol.* DOI
308 10.1111/2041-210X.12489.
- 309 Austerlitz, F., B. Jung-Muller, B. Godelle, and P.-H. Gouyon, 1997. Evolution of
310 coalescence times, genetic diversity and structure during colonization. *Theor.*
311 *Popul. Biol.* 51:148–164.
- 312 Barton, N. H. and G. M. Hewitt, 1985. Analysis of hybrid zones. *Annu. Rev.*
313 *Ecol. Syst.* 16:113–148.
- 314 Becquet, C. and M. Przeworski, 2009. Learning about modes of speciation by
315 computational approaches. *Evolution* 63:2547–2562.
- 316 Bermond, G., M. Ciosi, E. Lombaert, A. Blin, M. Boriani, L. Furian, S. Toepfer,
317 and T. Guillemaud, 2012. Secondary contact and admixture between indepen-
318 dently invading populations of the western corn rootworm, *Diabrotica virgifera*
319 *virgifera* in Europe. *PLoS One* 7:e50129.
- 320 Blair, C., D. E. Weigel, M. Balazik, A. T. Keeley, F. M. Walker, E. Landguth,
321 S. Cushman, M. Murphy, L. Waits, and N. Balkenhol, 2012. A simulation-
322 based evaluation of methods for inferring linear barriers to gene flow. *Mol.*
323 *Ecol. Resour.* 12:822–833.
- 324 Bray, S. M., J. G. Mulle, A. F. Dodd, A. E. Pulver, S. Wooding, and S. T.
325 Warren, 2010. Signatures of founder effects, admixture, and selection in the
326 Ashkenazi Jewish population. *Proc. Natl. Acad. Sci. U.S.A.* 107:16222–16227.

- 327 Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams,
328 A. Froment, J.-M. Bodo, C. Wambebe, S. A. Tishkoff, et al., 2010. Genome-
329 wide patterns of population structure and admixture in West Africans and
330 African Americans. *Proc. Natl. Acad. Sci. U.S.A.* 107:786–791.
- 331 Chakraborty, R. and K. M. Weiss, 1988. Admixture as a tool for finding linked
332 genes and detecting that difference from allelic association between loci. *Proc.*
333 *Natl. Acad. Sci. U.S.A.* 85:9119–9123.
- 334 Duforet-Frebourg, N. and M. G. B. Blum, 2014. Non-stationary patterns of
335 isolation-by-distance: inferring measures of local genetic differentiation with
336 Bayesian kriging. *Evolution* 68:1110–1123.
- 337 Durand, E., F. Jay, O. E. Gaggiotti, and O. François, 2009. Spatial inference of
338 admixture proportions and secondary contact zones. *Mol. Biol. Evol.* 29:1963–
339 1973.
- 340 Duvaux, L., K. Belkhir, M. Boulesteix, and P. Boursot, 2011. Isolation and
341 gene flow: inferring the speciation history of European house mice. *Mol.*
342 *Ecol.* 20:5248–5264.
- 343 Endler, J. A., 1977. *Geographic Variation, Speciation, and Clines*. Princeton
344 University Press.
- 345 Engelhardt, B. E. and M. Stephens, 2010. Analysis of population structure: a
346 unifying framework and novel methods based on sparse factor analysis. *PLoS*
347 *Genet.* 6:e1001117.
- 348 Frantz, A., S. Cellina, A. Krier, L. Schley, and T. Burke, 2009. Using spa-
349 tial Bayesian methods to determine the genetic structure of a continuously
350 distributed population: clusters or isolation by distance? *J. Appl. Ecol.*
351 46:493–505.

- 352 Hardy, O. J. and X. Vekemans, 1999. Isolation by distance in a continuous popu-
353 lation: reconciliation between spatial autocorrelation analysis and population
354 genetics models. *Heredity* 86:145–154.
- 355 Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–
356 913.
- 357 Hewitt, G. M., 1999. Post-glacial re-colonization of European biota. *Biol. J.*
358 *Linn. Soc.* 68:87–112.
- 359 Huber, C. D., M. Nordborg, J. Hermisson, and I. Hellmann, 2014. Keeping it
360 local: evidence for positive selection in Swedish *Arabidopsis thaliana*. *Mol.*
361 *Biol. Evol.* 31:3026–3039.
- 362 Hudson, R. R., 2002. Generating samples under a Wright-Fisher neutral model
363 of genetic variation. *Bioinformatics* 18:337–338.
- 364 Kelleher, J., N. H. Barton, and A. M. Etheridge, 2013. Coalescent simulation
365 in continuous space. *Bioinformatics* 29:955–956.
- 366 Keller, S. and D. Taylor, 2010. Genomic admixture increases fitness during a
367 biological invasion. *J. Evolution. Biol.* 23:1720–1731.
- 368 Kindler, E., R. Arlettaz, and G. Heckel, 2012. Deep phylogeographic diver-
369 gence and cytonuclear discordance in the grasshopper *Oedaleus decorus*. *Mol.*
370 *Phylogenet. Evol.* 65:695–704.
- 371 Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet, 2003. Landscape
372 genetics: combining landscape ecology and population genetics. *Trends Ecol.*
373 *Evol.* 18:189–197.
- 374 McVean, G. A. T., 2002. A genealogical interpretation of linkage disequilibrium.
375 *Genetics* 162:987–991.

- 376 Meirmans, P. G., 2012. The trouble with isolation by distance. *Mol. Ecol.*
377 21:2839–2846.
- 378 Murray, M. C. and M. P. Hare, 2006. A genomic scan for divergent selection
379 in a secondary contact zone between Atlantic and Gulf of Mexico oysters,
380 *Crassostrea virginica*. *Mol. Ecol.* 15:4229–4242.
- 381 Nielsen, E. E., M. M. Hansen, D. E. Ruzzante, D. Meldrup, and P. Grønkjær,
382 2003. Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic
383 and the Danish Belt Sea revealed by individual admixture analysis. *Mol. Ecol.*
384 12:1497–1508.
- 385 Novembre, J. and M. Stephens, 2008. Interpreting principal component analyses
386 of spatial population genetic variation. *Nat. Genet.* 10:646–649.
- 387 Paschou, P., E. Ziv, E. G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M. W.
388 Mahoney, and P. Drineas, 2007. PCA-correlated SNPs for structure identifi-
389 cation in worldwide human populations. *PLoS Genet.* 3:e160.
- 390 Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Gen-
391 schoreck, T. Webster, and D. Reich, 2012. Ancient admixture in human
392 history. *Genetics* 192:1065–1093.
- 393 Peter, B. M. and M. Slatkin, 2013. Detecting range expansions from genetic
394 data. *Evolution* 67:3274–3289.
- 395 Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Ched-
396 dadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, et al., 2003. Glacial
397 refugia: hotspots but not melting pots of genetic diversity. *Science* 300:1563–
398 1565.
- 399 Pinho, C. and J. Hey, 2010. Divergence with gene flow: models and data. *Annu.*
400 *Rev. Ecol. Evol. S.* 41:215–230.

- 401 Qu, Y., R. Zhang, Q. Quan, G. Song, S. H. Li, and F. Lei, 2012. Incomplete lin-
402 eage sorting or secondary admixture: disentangling historical divergence from
403 recent gene flow in the vinous-throated parrotbill (*Paradoxornis webbianus*).
404 Mol. Ecol. 21:6117–6133.
- 405 R Core Team, 2012. R: A Language and Environment for Statistical Computing.
406 R Foundation for Statistical Computing, Vienna, Austria. URL [http://www.](http://www.R-project.org/)
407 [R-project.org/](http://www.R-project.org/). ISBN 3-900051-07-0.
- 408 Ray, N., M. Currat, and L. Excoffier, 2010. SPLATCHE2: a spatially-explicit
409 simulation framework for complex demography, genetic admixture and recom-
410 bination. Bioinformatics 26:2993–2994. URL <http://www.splatche.com/>.
- 411 Riley, S. P., J. P. Pollinger, R. M. Sauvajot, E. C. York, C. Bromley, T. K. Fuller,
412 and R. K. Wayne, 2006. FAST-TRACK: A southern California freeway is a
413 physical and social barrier to gene flow in carnivores. Mol. Ecol. 15:1733–1741.
- 414 Roux, C., G. Tsagkogeorga, N. Bierne, and N. Galtier, 2013. Crossing the species
415 barrier: genomic hotspots of introgression between two highly divergent *Ciona*
416 *intestinalis* species. Mol. Biol. Evol. 30:1574–1587.
- 417 Safner, T., M. P. Miller, B. H. McRae, M.-J. Fortin, and S. Manel, 2011.
418 Comparison of Bayesian clustering and edge detection methods for inferring
419 boundaries in landscape genetics. Int. J. Mol. Sci. 12:865–889.
- 420 Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger, and L. P. Waits, 2010.
421 Landscape genetics: where are we now? Mol. Ecol. 19:3496–3514.
- 422 Strand, A. E., L. M. Williams, M. F. Oleksiak, and E. E. Sotka, 2012. Can diver-
423 sifying selection be distinguished from history in geographic clines? A popu-
424 lation genomic study of killifish (*Fundulus heteroclitus*). PLoS One 7:e45138.

425 Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy, and J.-F. Cosson, 1998. Com-
426 parative phylogeography and postglacial colonization routes in Europe. *Mol.*
427 *Ecol.* 7:453–464.

428 Zalewski, A., S. B. Piertney, H. Zalewska, and X. Lambin, 2009. Landscape
429 barriers reduce gene flow in an invasive carnivore: geographical and local
430 genetic structure of American mink in Scotland. *Mol. Ecol.* 18:1601–1615.

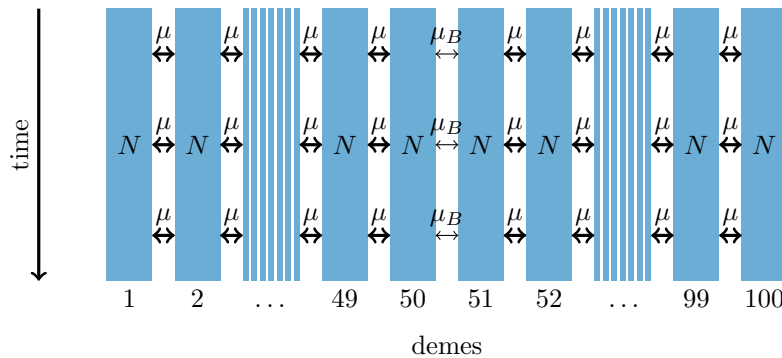


Figure 2: 1-dimensional model of a barrier to gene flow in a nearest-neighbor stepping-stone environment with 100 demes.

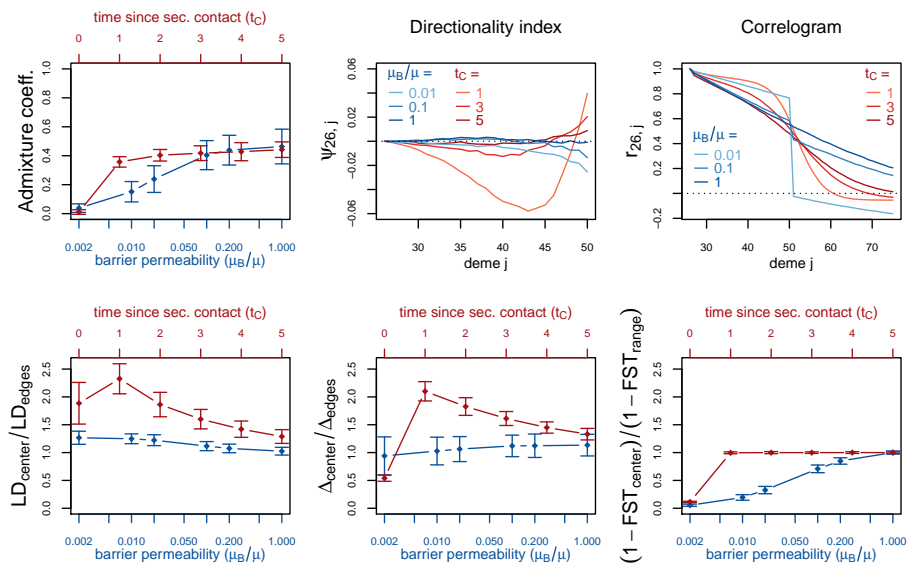


Figure 3: Red (lines and axes): secondary contact model; blue: stepping-stone model with barrier. The barrier permeability μ_B/μ is plotted on a logarithmic scale with $\mu = 20$. The dots denote the mean and the error bars ± 2 standard errors, estimated from 100 replicates of the simulations. For Δ and LD , the subscript *center* denotes the mean over demes 49–52 and *edges* over demes 24–27 and 74–77. Here, *center* denotes demes 50 and 51 and *range* the mean over the neighboring demes in 26–74 except demes 50 and 51. (For these statistics, the edges of the range are dismissed because of the edge-effect in the stepping-stone model.) For the allele frequency correlogram and the ψ statistic, only the mean is plotted.

t_C	Secondary contact		μ_B/μ	Barrier	
	Spearman	$p < 0.05$		Spearman	$p < 0.05$
0	0.16	6	0.002	0.05	62
1	0.53	100	0.01	0.10	71
2	0.58	100	0.02	0.13	84
3	0.60	100	0.1	0.18	87
4	0.58	100	0.2	0.19	97
5	0.58	100	1	0.21	94

Table 1: Test for admixture based on LD. Left: Secondary contact model with different times since secondary contact (t_C); right: stepping-stone model with barrier to gene flow of different intensities (μ_B/μ ; $\mu = 20$). The mean Spearman correlation coefficient over 100 replicates of the simulations. The column $p < 0.05$ gives the number replicates with coefficients that are significantly larger than 0.

Supplementary Information

1 Examples of summary statistics along the range of demes

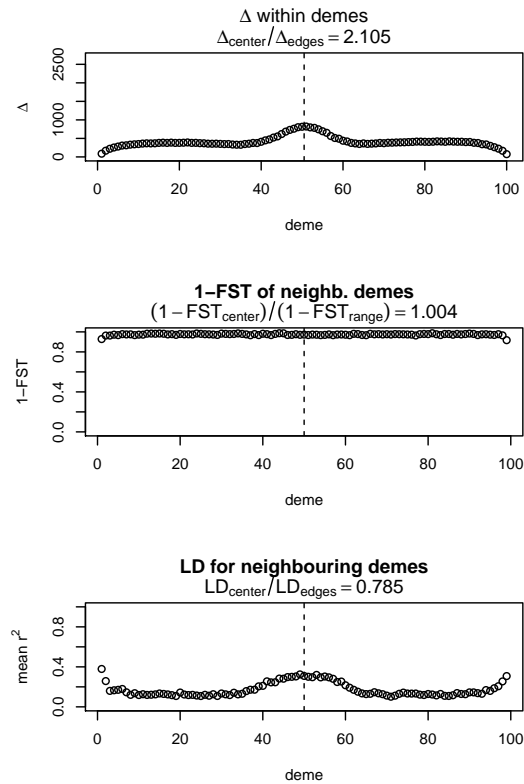


Figure S1: Simulation results for 1 replicate of the secondary contact model with $t_C = 1$.

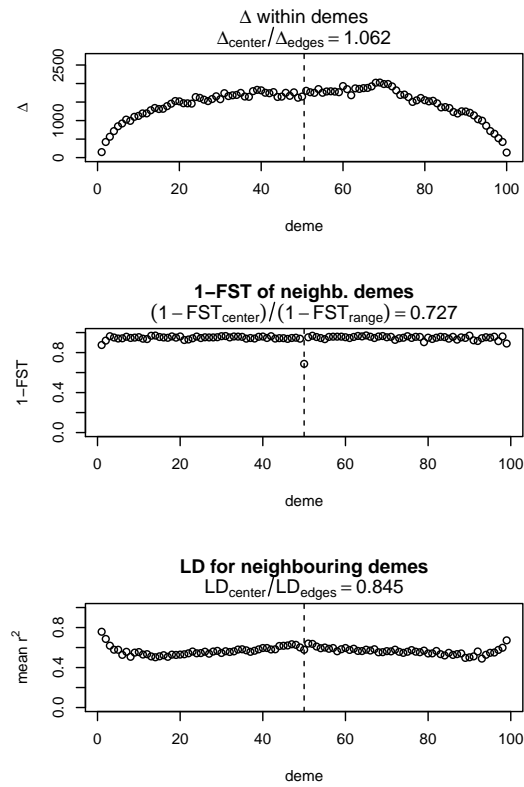


Figure S2: Simulation results for 1 replicate of the stepping stone model with $\mu_B = 2$.

2 Simulations at additional parameter combinations

To assess the sensitivity of the results with regard to the other parameters of the models, we resimulated under the secondary contact model with $t_C = 1$ and varied the bottleneck intensity (fig. S3) and the expansion speed (d ; fig S4). For comparison, the stepping stone model is shown with the same parameters as in the main text.

In addition, we vary the mutation rate μ simultaneously in both models, (fig. S5), in the secondary contact model with $t_C = 1$ and in the stepping stone model with $\mu_B/\mu = 0.002$.

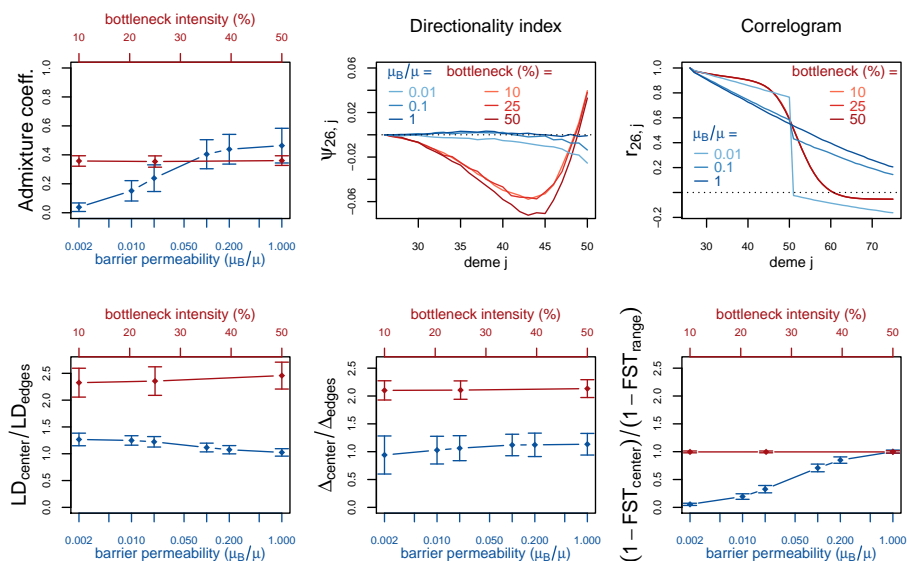


Figure S3: Red (lines and axis): secondary contact model with $t_C = 1$ and varying bottleneck intensity. Blue (lines and axis): stepping stone model with barrier. The dots denote the mean and the error bars ± 2 standard errors, estimated from 100 replicates of the simulations. For Δ and LD , the subscript *center* denotes the mean over demes 49–52 and *edges* over demes 24–27 and 74–77. Here, *center* denotes demes 50 and 51 and *range* the mean over the neighboring demes in 26–74 except demes 50 and 51. (For these statistics, the edges of the range are dismissed because of the edge-effect in the stepping-stone model.) For the allele frequency correlogram and the ψ statistic, only the mean is plotted. Note that the correlogram is so similar for each of the three parameters that they are plotted on top of each other.

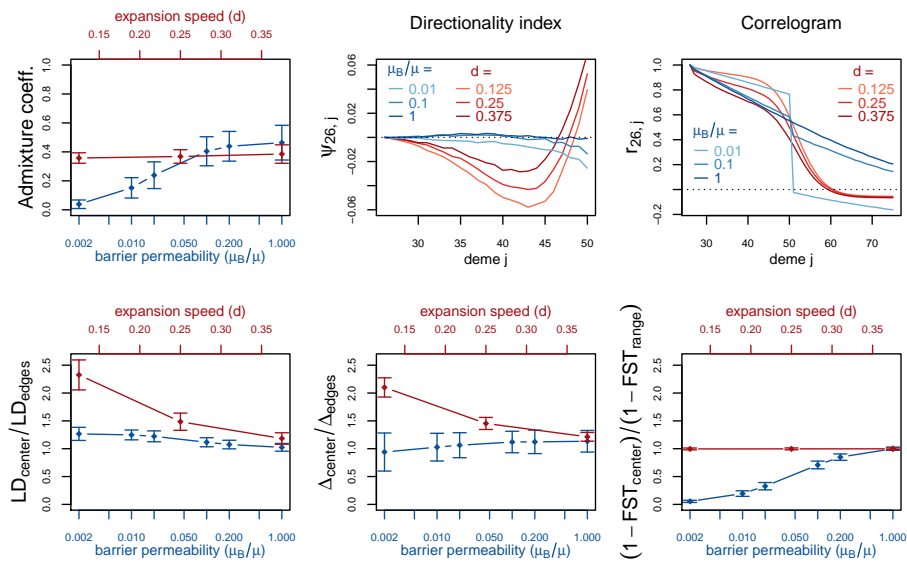


Figure S4: Red (lines and axis): secondary contact model with $t_C = 1$ and varying expansion speed (d). Blue (lines and axis): stepping stone model with barrier. Definitions of the summary statistics as in fig. S3.

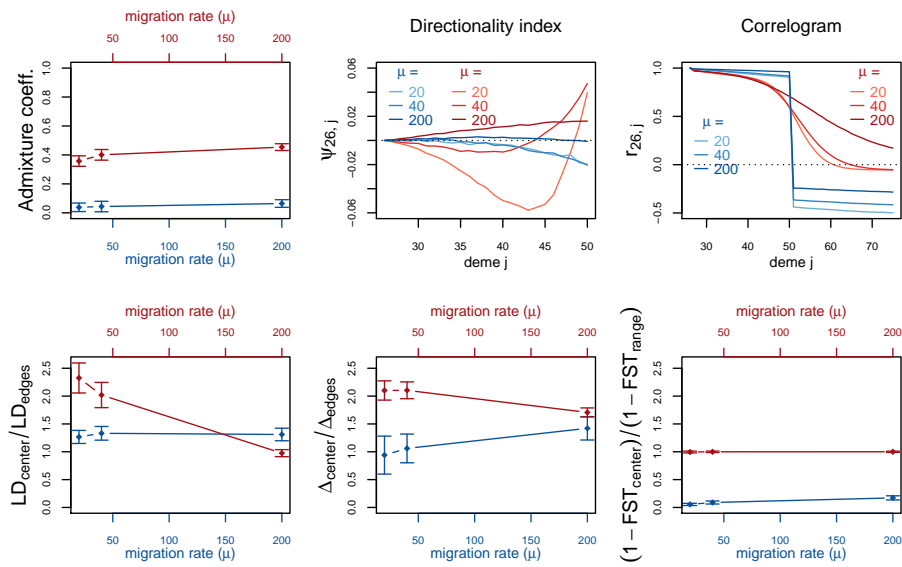


Figure S5: Red (lines and axis): secondary contact model with $t_C = 1$ and varying migration rate (μ). Blue (lines and axis): stepping stone model with varying migration rate (μ). The barrier permeability $\mu_B/\mu = 0.002$ is constant. Definitions of the summary statistics as in fig. S3.

3 Secondary contact model with barrier

We also simulate under a secondary contact model with a moderate barrier to gene flow at the secondary contact zone ($\mu_B = 2$, $\mu = 20$). For comparison, the secondary contact zone without barrier and the stepping stone model from fig. 1 in the main text are plotted again.

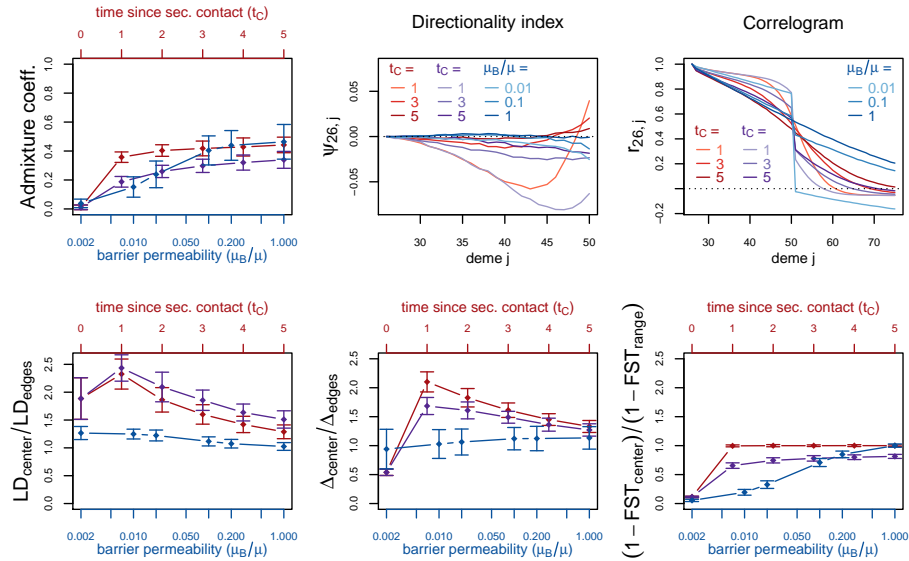


Figure S6: Red (lines and axes): secondary contact model; blue: stepping-stone model; both models as in fig. 1 (main text). Purple: secondary contact model with a moderate barrier to gene flow ($\mu_B/\mu = 0.1$) at the secondary contact zone (refer to the red axis for the values of t_C). Definitions of the summary statistics as in fig. S3.