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

Johanna Bertl[1] 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

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

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

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

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

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

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

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

2 Methods

2.1 Models

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

with $4N$.

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

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

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

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

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

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

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

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

2.2 Summary statistics

Hybrid zone summary statistics

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

The principal component analysis is conducted with the `R` function `prcomp` (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

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

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

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

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

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

Range expansion summary statistics

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

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

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

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

Local $\mathbf{F_{ST}}$ Local values of F_{ST} correspond to pairwise F_{ST} between neighboring demes (Duforet-Frebourg and Blum, 2014).

3 Results

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

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

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

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

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

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

diversity when moving away from the origin.

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

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

4 Discussion

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

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

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

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

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

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

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

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

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

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

Acknowledgments

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

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

References

Adams, S. M., J. B. Lindmeier, and D. D. Duvernell, 2006. Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. Mol. Ecol. 15:1109–1123.

Alvarado-Serrano, D. F. and M. J. Hickerson, 2015. Spatially explicit summary statistics for historical population genetic inference. Methods Ecol. Evol. DOI 10.1111/2041-210X.12489.

Austerlitz, F., B. Jung-Muller, B. Godelle, and P.-H. Gouyon, 1997. Evolution of coalescence times, genetic diversity and structure during colonization. Theor. Popul. Biol. 51:148–164.

Barton, N. H. and G. M. Hewitt, 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Becquet, C. and M. Przeworski, 2009. Learning about modes of speciation by computational approaches. Evolution 63:2547–2562.

Bermond, G., M. Ciosi, E. Lombaert, A. Blin, M. Boriani, L. Furian, S. Toepfer, and T. Guillemaud, 2012. Secondary contact and admixture between independently invading populations of the western corn rootworm, *Diabrotica virgifera virgifera* in Europe. PLoS One 7:e50129.

Blair, C., D. E. Weigel, M. Balazik, A. T. Keeley, F. M. Walker, E. Landguth, S. Cushman, M. Murphy, L. Waits, and N. Balkenhol, 2012. A simulation-based evaluation of methods for inferring linear barriers to gene flow. Mol. Ecol. Resour. 12:822–833.

Bray, S. M., J. G. Mulle, A. F. Dodd, A. E. Pulver, S. Wooding, and S. T. Warren, 2010. Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. Proc. Natl. Acad. Sci. U.S.A. 107:16222–16227.

Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams, A. Froment, J.-M. Bodo, C. Wambebe, S. A. Tishkoff, et al., 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. U.S.A. 107:786–791.

Chakraborty, R. and K. M. Weiss, 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc. Natl. Acad. Sci. U.S.A. 85:9119–9123.

Duforet-Frebourg, N. and M. G. B. Blum, 2014. Non-stationary patterns of isolation-by-distance: inferring measures of local genetic differentiation with Bayesian kriging. Evolution 68:1110–1123.

Durand, E., F. Jay, O. E. Gaggiotti, and O. François, 2009. Spatial inference of admixture proportions and secondary contact zones. Mol. Biol. Evol. 29:1963–1973.

Duvaux, L., K. Belkhir, M. Boulesteix, and P. Boursot, 2011. Isolation and gene flow: inferring the speciation history of European house mice. Mol. Ecol. 20:5248–5264.

Endler, J. A., 1977. Geographic Variation, Speciation, and Clines. Princeton University Press.

Engelhardt, B. E. and M. Stephens, 2010. Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis. PLoS Genet. 6:e1001117.

Frantz, A., S. Cellina, A. Krier, L. Schley, and T. Burke, 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? J. Appl. Ecol. 46:493–505.

Hardy, O. J. and X. Vekemans, 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 86:145–154.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.

Hewitt, G. M., 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68:87–112.

Huber, C. D., M. Nordborg, J. Hermisson, and I. Hellmann, 2014. Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana*. Mol. Biol. Evol. 31:3026–3039.

Hudson, R. R., 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Kelleher, J., N. H. Barton, and A. M. Etheridge, 2013. Coalescent simulation in continuous space. Bioinformatics 29:955–956.

Keller, S. and D. Taylor, 2010. Genomic admixture increases fitness during a biological invasion. J. Evolution. Biol. 23:1720–1731.

Kindler, E., R. Arlettaz, and G. Heckel, 2012. Deep phylogeographic divergence and cytonuclear discordance in the grasshopper *Oedaleus decorus*. Mol. Phylogenet. Evol. 65:695–704.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet, 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189–197.

McVean, G. A. T., 2002. A genealogical interpretation of linkage disequilibrium. Genetics 162:987–991.

Meirmans, P. G., 2012. The trouble with isolation by distance. Mol. Ecol. 21:2839–2846.

Murray, M. C. and M. P. Hare, 2006. A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. Mol. Ecol. 15:4229–4242.

Nielsen, E. E., M. M. Hansen, D. E. Ruzzante, D. Meldrup, and P. Grønkjær, 2003. Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. Mol. Ecol. 12:1497–1508.

Novembre, J. and M. Stephens, 2008. Interpreting principal component analyses of spatial population genetic variation. Nat. Genet. 10:646–649.

Paschou, P., E. Ziv, E. G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M. W. Mahoney, and P. Drineas, 2007. PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genet. 3:e160.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich, 2012. Ancient admixture in human history. Genetics 192:1065–1093.

Peter, B. M. and M. Slatkin, 2013. Detecting range expansions from genetic data. Evolution 67:3274–3289.

Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, et al., 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300:1563–1565.

Pinho, C. and J. Hey, 2010. Divergence with gene flow: models and data. Annu. Rev. Ecol. Evol. S. 41:215–230.

Qu, Y., R. Zhang, Q. Quan, G. Song, S. H. Li, and F. Lei, 2012. Incomplete lineage sorting or secondary admixture: disentangling historical divergence from recent gene flow in the vinous-throated parrotbill (*Paradoxornis webbianus*). Mol. Ecol. 21:6117–6133.

R Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

Ray, N., M. Currat, and L. Excoffier, 2010. SPLATCHE2: a spatially-explicit simulation framework for complex demography, genetic admixture and recombination. Bioinformatics 26:2993–2994. URL `http://www.splatche.com/`.

Riley, S. P., J. P. Pollinger, R. M. Sauvajot, E. C. York, C. Bromley, T. K. Fuller, and R. K. Wayne, 2006. FAST-TRACK: A southern California freeway is a physical and social barrier to gene flow in carnivores. Mol. Ecol. 15:1733–1741.

Roux, C., G. Tsagkogeorga, N. Bierne, and N. Galtier, 2013. Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. Mol. Biol. Evol. 30:1574–1587.

Safner, T., M. P. Miller, B. H. McRae, M.-J. Fortin, and S. Manel, 2011. Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. Int. J. Mol. Sci. 12:865–889.

Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger, and L. P. Waits, 2010. Landscape genetics: where are we now? Mol. Ecol. 19:3496–3514.

Strand, A. E., L. M. Williams, M. F. Oleksiak, and E. E. Sotka, 2012. Can diversifying selection be distinguished from history in geographic clines? A population genomic study of killifish (*Fundulus heteroclitus*). PLoS One 7:e45138.

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

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

Figures and tables

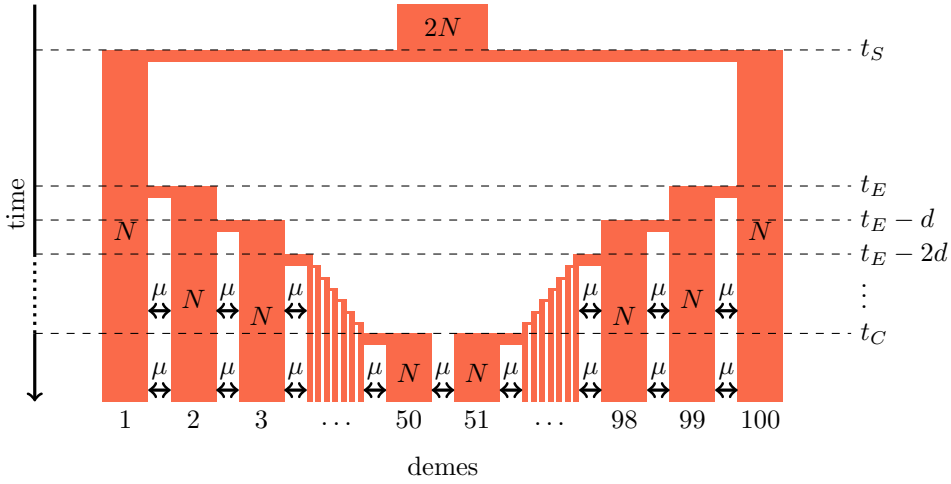

Figure 1: 1-dimensional secondary contact model in a nearest-neighbor stepping-stone environment with 100 demes.

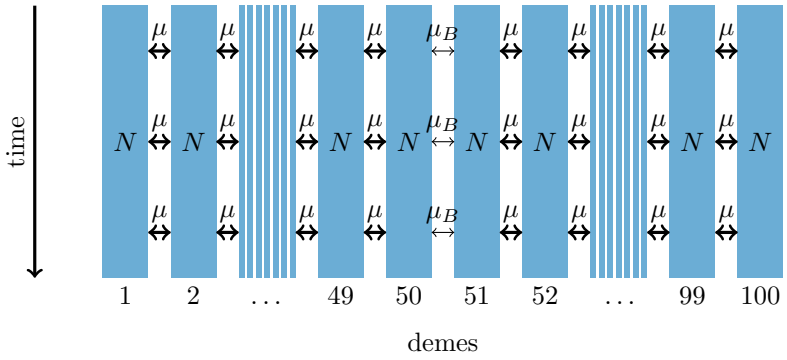

Figure 2: 1-dimensional model of a barrier to gene flow in a nearest-neighbor stepping-stone environment with 100 demes.

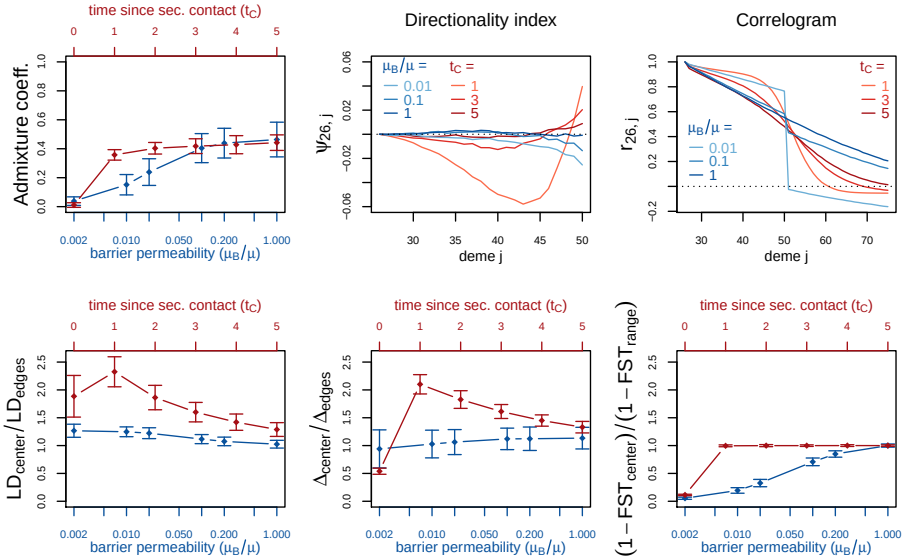

Figure 3: Red (lines and axes): secondary contact model; blue: stepping-stone model with barrier. The barrier permeability μ_B/μ is plotted on a logarithmic scale with $\mu = 20$. The dots denote the mean and the error bars ± 2 standard errors, estimated from 100 replicates of the simulations. For Δ 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.

Secondary contact			Barrier		
t_C	Spearman	$p < 0.05$	μ_B/μ	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

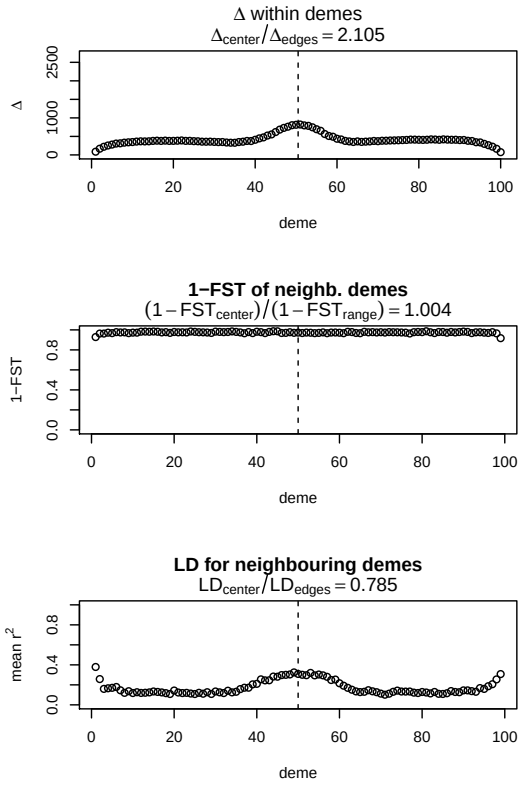

Figure S1: Simulation results for 1 replicate of the secondary contact model with $t_C = 1$.

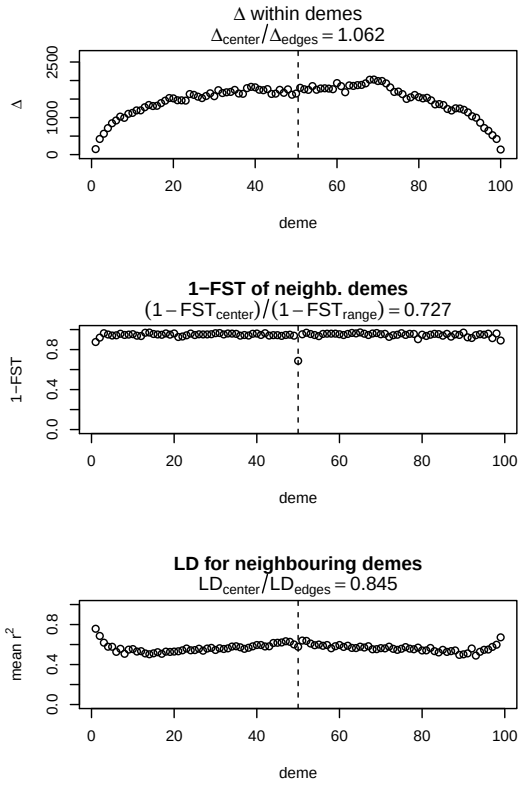

Figure S2: Simulation results for 1 replicate of the stepping stone model with $\mu_B = 2$.

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 $\mu_B/\mu = 0.002$.

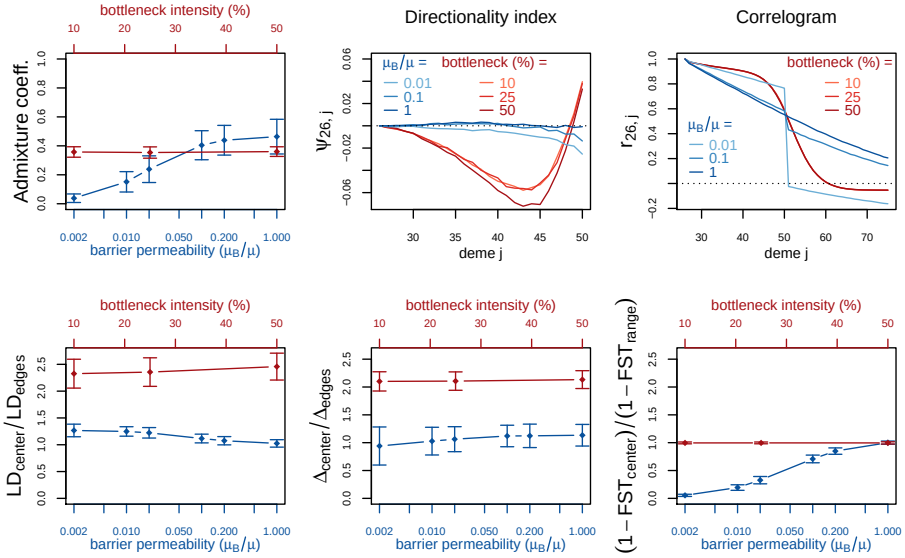

Figure S3: Red (lines and axis): secondary contact model with $t_C = 1$ and varying bottleneck intensity. Blue (lines and axis): stepping stone model with barrier. The dots denote the mean and the error bars ±2 standard errors, estimated from 100 replicates of the simulations. For Δ and LD, the subscript *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.

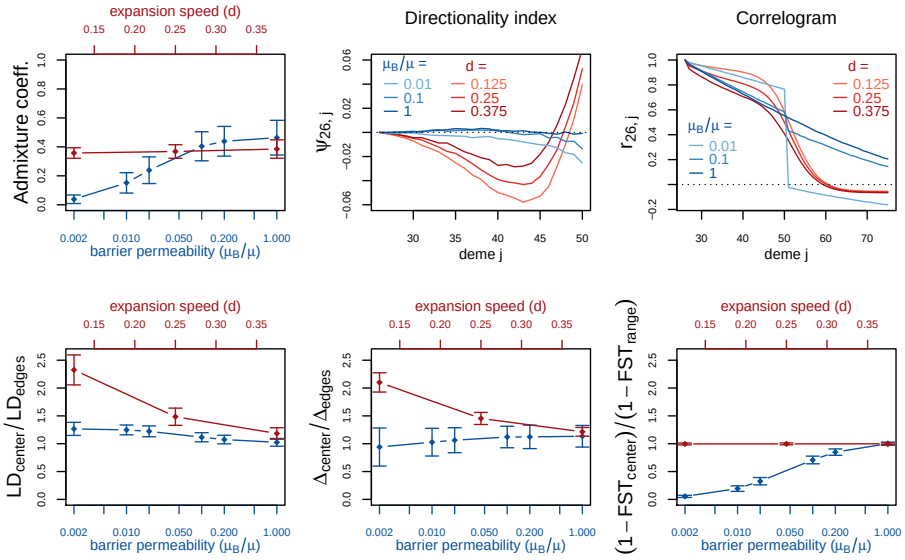

Figure S4: Red (lines and axis): secondary contact model with $t_C = 1$ and varying expansion speed (d). Blue (lines and axis): stepping stone model with barrier. Definitions of the summary statistics as in fig. S3.

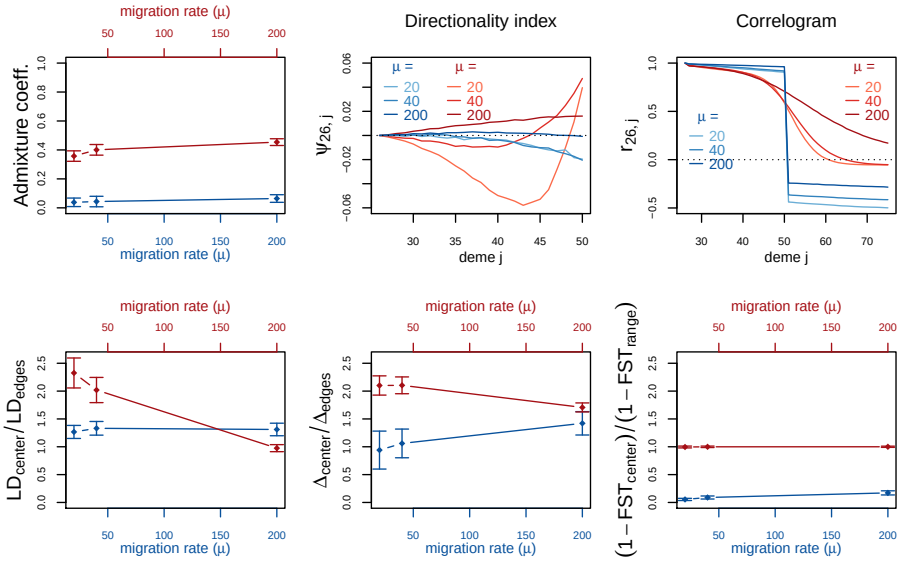

Figure S5: Red (lines and axis): secondary contact model with $t_C = 1$ and varying migration rate (μ). Blue (lines and axis): stepping stone model with varying migration rate(μ). The barrier permeability $\mu_B/\mu = 0.002$ is constant. Definitions of the summary statistics as in fig. S3.

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.

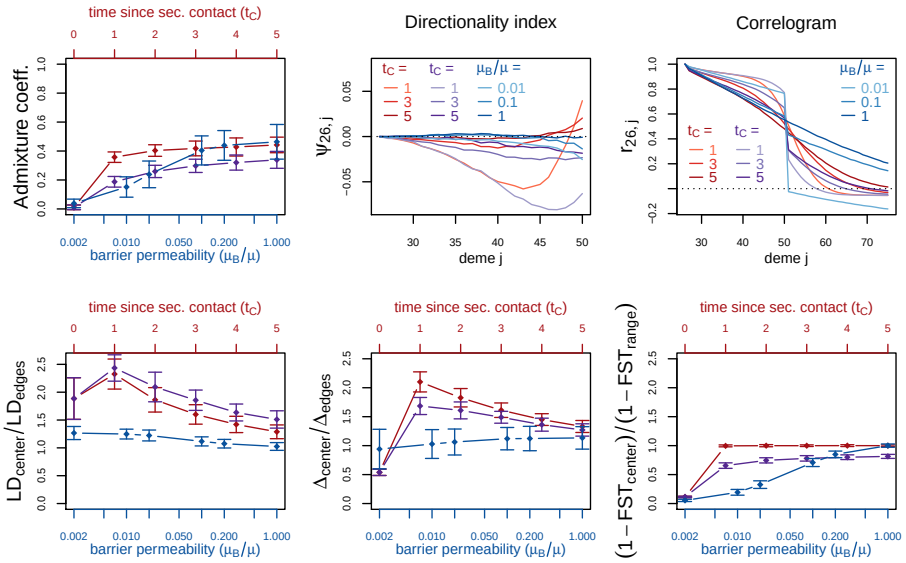

Figure S6: Red (lines and axes): secondary contact model; blue: stepping-stone model; both models as in fig. 1 (main text). Purple: secondary contact model with a moderate barrier to gene flow ($\mu_B/\mu = 0.1$) at the secondary contact zone (refer to the red axis for the values of t_C). Definitions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