



HAL
open science

The geography of introgression in a patchy environment and the thorn in the side of ecological speciation

Nicolas Bierne, Pierre-Alexandre Gagnaire, Patrice David

► **To cite this version:**

Nicolas Bierne, Pierre-Alexandre Gagnaire, Patrice David. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current zoology*, 2013, 59 (1), pp.72 - 86. 10.1093/czoolo/59.1.72 . hal-02395293

HAL Id: hal-02395293

<https://hal.science/hal-02395293v1>

Submitted on 5 Dec 2019

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.

The geography of introgression in a patchy environment and the thorn in the side of ecological speciation

Nicolas BIERNE^{1,2*}, Pierre-Alexandre GAGNAIRE^{1,2}, Patrice DAVID^{1,3}

¹ Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

² ISEM - CNRS, UMR 5554, SMEL, 2 rue des Chantiers, 34200 Sète, France

³ CEFE - CNRS, UMR 5175, 34293 Montpellier Cedex 5, France

Abstract When incompletely isolated taxa coexist in a patchy environment (e.g. mosaic hybrid zones, host-race complexes), patterns of variation may differ between selected traits/genes and neutral markers. While the genetic structure of selected traits/loci tends to coincide with habitat variables (producing Genetic-Environment Association or GEA), genetic differentiation at neutral loci unlinked to any selected locus rather depends on geographic connectivity at a large scale (e.g. Isolation-By-Distance or IBD), although these loci often display GEA at a small scale. This discrepancy has been repeatedly taken as evidence for parallel primary divergence driven by local adaptation. We argue that this interpretation needs to be addressed more thoroughly by considering the alternative hypothesis that speciation was initiated in allopatry and secondary introgression has subsequently erased the signal of past differentiation at neutral loci. We present a model of neutral introgression after secondary contact in a mosaic hybrid zone, which describes how GEAs dissipate with time and how neutral variation self-organizes according to the environmental and geographic structures. We show that although neutral loci can be affected by environmental selection, they are often more affected by history and connectivity: the neutral structure retains the initial geographic separation more than it correlates with the environment during the colonization and introgression phases, and then converges to a migration-drift balance, the most frequent outcome of which is GEA at a local scale but IBD at a large scale. This is the exact pattern usually attributed to parallel ecological speciation. Introgression is heterogeneous in space and depends on the landscape structure (e.g. it is faster in small patches, which are more impacted by immigration). Furthermore, there is no directionality in the association and it is possible to observe reversed GEAs between distant regions. We argue that the history of differentiation should ideally be reconstructed with selected loci or neutral loci linked to them, not neutral ones, and review some case studies for which the hypothesis of a long co-existence of co-adapted genetic backgrounds might have been refuted too hastily [*Current Zoology* 59 (1): 72–86, 2013].

Keywords Ecological speciation, Local adaptation, Mosaic hybrid zone, Reproductive isolation, Genetic-environment association

Speciation is usually defined as an evolutionary process during which diverging populations accumulate reproductive barriers until they become completely isolated (Coyne and Orr, 2004). Several mechanisms exist that account for the evolution of reproductive barriers in the presence of gene flow, with initial conditions ranging from primary divergence in sympatry to secondary contact after geographic isolation (Endler, 1977; Smadja and Butlin, 2011). In natural populations, speciation is often studied in admixed populations or hybrid zones, which offer ideal conditions for studying the flow of alleles between partially isolated backgrounds (Barton and Hewitt, 1985). However, it has long been recognized how difficult it is to discriminate secondary contact zones from ongoing primary differentiation of parapatrically isolated forms (primary *versus* secondary intergradation, Endler, 1977; Harrison, 2011). The rea-

son is that, in secondary contact, neutral markers reveal the history of geographic isolation only transiently and converge to an equilibrium which is the same as for primary differentiation –i.e. two parapatric forms isolated by a semi-permeable genetic barrier to gene flow (Barton and Hewitt, 1985; Harrison, 1993; Charlesworth et al., 1997). Thus, although the initial conditions are different between the two scenarios, they eventually converge to the same state (Endler, 1977; Harrison, 2012).

While the introgression process is well understood in the case of a single genetic barrier separating two parapatric forms (Endler, 1977; Barton and Bengtsson, 1986; Charlesworth et al., 1997), it has rarely been described when the spatial pattern is patchy (e.g. mosaic hybrid zones, host races complexes, pairs of ecotypes in multiple sympatric locations). This situation occurs when the environment consists of a mosaic of habitat patches that

Received July 26, 2012; accepted Nov. 26, 2012.

* Corresponding author. E-mail: n-bierne@univ-montp2.fr

© 2013 *Current Zoology*

fall into two types, to which two taxa (ecotypes, races, subspecies or hybridizing species) are somehow specialized. Neutral alleles (e.g. microsatellites) typical of each taxon are initially associated with their respective preferred habitat (Genetic-Environment Associations or GEAs), and it is often assumed that when both taxa come into secondary contact, neutral loci will retain this pattern for a long time, just the same way as neutral differentiation would persist across a “single-cline” tension zone. Consequently, when GEA is observed only at a few selected traits/loci, while at the same time differentiation at neutral loci rather reflects geographic distance (Isolation-By-distance or IBD), the hypothesis of

secondary contact is usually put aside and data are interpreted as evidence for a primary divergence driven by local adaptation – i.e. that selection is currently acting to increase differentiation between habitats, at different geographical locations, in a parallel fashion (ongoing, parallel ecological speciation). Such conflicting patterns between neutral and selected markers are illustrated in Figure 1 and have been repeatedly observed in population genomic studies of ecologically divergent species pairs (e.g. Wilding et al., 2001; Campbell and Bernatchez, 2004; Colosimo et al., 2005; Bonin et al., 2006; Egan et al., 2008; Nosil et al., 2008; Via and West, 2008).

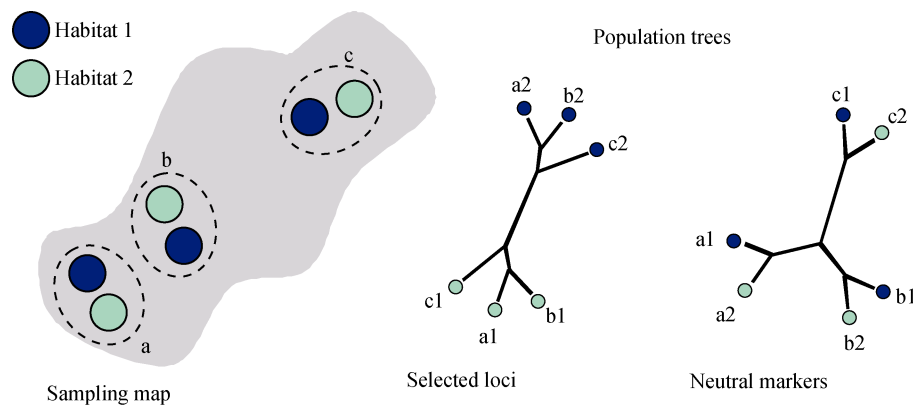


Fig. 1 The thorn in the side of ecological speciation

When incompletely isolated taxa coexist in a patchy environment, patterns of variation may differ between selected and neutral loci. The genetic structure at selected loci coincides with habitat variation. Neutral loci unlinked to any selected locus rather reflect geographic connectivity although these loci can still reveal genetic differentiation between habitats locally. The genetic structure is here presented with population trees, as it is commonly done. Population trees inform on the genetic similarities between population samples, but by no ways should be interpreted as revealing the history of populations (as a phylogenetic tree could do). Genetic similarities can result from a shared history but also gene flow or selection to a similar environment. A conflictual pattern between neutral and selected loci is often interpreted as evidence for ongoing primary differentiation while it is expected under a scenario of secondary contact as well.

Our aim here is to illustrate that neutral markers can easily and quickly lose most of their associations with selected loci, and hence with the environment, under both primary and secondary intergradation. Using a simple model of a mosaic hybrid zone, we describe how GEAs at neutral loci dissipate during secondary introgression depending on the initial conditions and the landscape structure, and how fast the system converges to the migration/drift equilibrium. We show that, after secondary contact, GEAs can be observed at neutral loci at a local geographic scale but that IBD is more frequent at a large scale. This is the exact pattern that is, perhaps erroneously, usually attributed to parallel adaptation in many popular models of the ecological speciation literature. The random distribution of habitat patches we model may well resemble the landscape faced by popular examples of ecological speciation (e.g. leaf beetles:

Egan et al., 2008; walking sticks: Nosil et al., 2008; pea aphids: Via and West, 2008). We further analyze two specific landscapes which illustrate two keystone case studies of the ecological speciation literature, the high shore/low shore landscapes of *Littorina* periwinkle ecotypes (e.g. Wilding et al., 2001; Panova et al., 2006; Quesada et al., 2007; Rolán-Alvarez, 2007; Galindo et al., 2009) and the marine/freshwater structure of three-spined sticklebacks (e.g. Colosimo, 2005; Hohenlohe et al., 2010; Deagle et al., 2011; Jones et al., 2012a). Contrary to a previous work (Bierne et al., 2011) our argument here is independent from the type of selection that acts on the barrier loci that prevent free interbreeding between taxa. Our argument requires that adaptation to habitat (exogenous selection) exists somewhere in the genome – otherwise GEA would be absent. Endogenous selection (habitat-independent selection due to intrinsic

genetic incompatibilities) may or may not also contribute to the genetic barrier. Barrier loci can also be involved in pre-zygotic isolation or habitat choice, although our model uses post-zygotic selection as it is straightforward to compute. We distinguish two types of neutral loci, those that are unlinked to any barrier loci (neutral markers *sensu stricto*) and those that are linked to a barrier locus. Neutral loci linked to barrier loci have a reduced effective migration rate and are here called genomic island (GI) loci (see Glossary), illustrating that they are in the chromosomal region of elevated differentiation around barrier loci. Because the language of speciation is complex (Harrison, 2012), with the same terms used with different meanings in different publications, we provide a glossary with the main definitions we used at the end of the article.

We argue that the contemporary distribution of taxa reveals little about the true spatial context in which the polymorphisms underlying adaptation to different habitats originally evolved, and that studies of ecological speciation have tended to neglect long term evolution by focusing on the most recent events, while most known genetic barriers are several glacial oscillations old.

1 Material and Method

1.1 General model

We used a model of evolution in a metapopulation of n demes arranged in a linear stepping-stone structure (see Bierne et al., 2011). In order to model a fine-grained environmental structure, the standard stepping-stone model in which only adjacent demes are connected by migration was modified so that migration not only connects adjacent demes but also demes separated by up to six demes (Fig. 2). The auto-recruitment rate is $1-m$; the migration rate to adjacent demes is $m/4$, to demes at a distance of two demes $m/8$, to demes at a distance of three demes $m/16$, at a distance of four demes $m/32$, and at a distance of five and six demes $m/64$ (m was set to 0.5 in the simulations presented here). Reflecting barriers are present at both ends of the chain of demes. Each deme in the chain was assigned to one of two habitats types (habitat 1 and habitat 2) randomly. We consider three bi-allelic haploid loci. The first two loci are barrier loci. They have alleles labelled A and a (locus 1) and B and b (locus 2), they are unlinked and are under selection. Locus 1 alleles are under exogenous selection, with allele A being adapted to habitat 1 and allele a to habitat 2. Locus 2 is involved in a symmetric genetic incompatibility with locus 1: B is incompatible with a , and b is incompatible with A . The

two-locus fitnesses are $W(AB)=1$, $W(Ab)=1-s$, $W(ab)=1-t$ and $W(aB)=(1-t)(1-s)$ in habitat 1, and $W(ab)=1$, $W(aB)=(1-s)$, $W(AB)=1-t$ and $W(Ab)=(1-t)(1-s)$ in habitat 2. The third locus is neutral, has alleles labeled C and c and is at a recombination distance r of locus 1 (however the same results are expected if locus 3 is linked to locus 2 as epistatic selection maintains associations between locus 1 and 2). Genotypic frequencies in each deme were derived from those of the previous generation after accounting for recombination, migration and selection in that order. Random genetic drift was simulated by multinomial sampling of genotypes within each deme at each generation. We consider two initial conditions. In the model called “instantaneous colonization” the simulation starts by considering each habitat to be fixed by the ABC genotype if it is habitat 1 and abc if it is habitat 2. In the model called “secondary contact” the left part of the chain of demes is initially fixed with ABC genotypes and the right part of the chain with abc , irrespective of the environment. Windows executable and Borland Delphi 4.0 source code are available from the authors upon request.

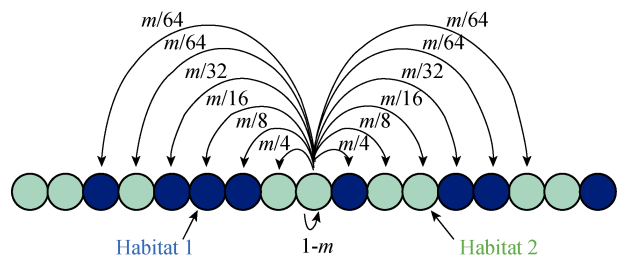


Fig. 2 Modified stepping stone model

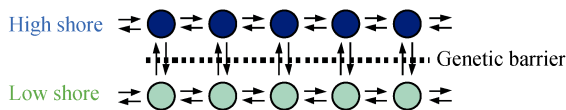
In order to model a fine-grained environmental structure, demes separated by up to six demes are connected by migration with a decreasing rate. Each deme is assigned a habitat type (blue: habitat 1, green: habitat 2) randomly.

1.2 Case studies

We have modeled two landscape structures that will serve to illustrate the environmental variation faced by the rough periwinkle *Littorina saxatilis* (Butlin et al., 2008; Johannesson et al., 2010) and the three-spined stickleback *Gasterosteus aculeatus* (Schluter and Conte, 2009). *L. saxatilis* inhabits the intertidal rocky shores throughout Europe. Two morphologically different ecotypes are usually found on the shore: a small wave-tolerant, thin-shelled morph with a wide aperture occupies the high shore, and a larger, thicker-shelled morph with a narrower aperture providing increased resistance to predation by crabs occupies the lower shore, forming hybrid zones in habitat transitions (Johannesson et al., 2010). To simulate this landscape structure we used two

parallel stepping stones, one corresponding to the high shore and the other to the low shore (Fig. 3A). The migration rate, m , between adjacent demes is the same, be they of the same or of different habitat type, and gene flow between habitats is reduced by indirect selection –*i.e.* a genetic barrier. To simulate the landscape faced by the marine and stream ecotypes of three-spined stickleback, we used a standard stepping stone to represent the sea population and individual demes connected to this marine stepping stone to represent streams (Fig. 3B). As suspected for this species, direct migration was not possible between streams (Schluter and Conte, 2009).

A. Case study 1: *Littorina periwinkles*



B. Case study 2: three-spined sticklebacks

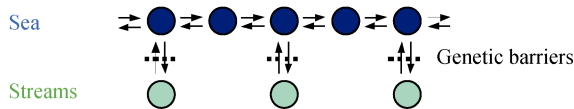


Fig. 3 Landscape structures used to model case studies

A. The rough periwinkle *Littorina saxatilis*. B. The three-spined stickleback *Gasterosteus aculeatus*.

2 Results

2.1 General model

At first sight the strength of a genetic barrier is expected to be roughly proportional to s/r (Barton 1979; Barton 1983; Barton and Bengtsson 1986). Here we are interested in genetic barriers with genome-wide effects that affect the effective migration rate of neutral loci unlinked to selected loci (a process recently coined “genome hitchhiking”, Feder et al., 2012), so that GEA can be observed with a handful of molecular markers (say microsatellites). We therefore need strong selection coefficients. Although it might appear more realistic to model many loci that conjugate their effects to form a strong barrier to interspecific gene flow, we used a more tractable model with just two loci, each with a strong selective coefficient ($s=t=0.9$; only 1% of recombinant genotypes survive in the wrong habitat). We used exogenous selection at one locus, and genetic incompatibility with a second locus, but similar results could have been obtained with exogenous selection only, providing a similar overall selection coefficient was used. The use of few loci is justified because when a barrier is strong

enough to impede the flow of unlinked neutral genes, it has reached a “congealed” state so that a collection of unlinked loci under concordant, moderate, selection would roughly behave as a single superlocus under strong selection because strong linkage disequilibrium is maintained between them (Barton 1983; Kruuk et al., 1999). In addition, we used post-zygotic selection to model the barrier, but any other type of isolation mechanism (e.g. assortative mating, habitat choice) is expected to produce similar effects (Gavrilets and Cruzan 1998).

In Fig. 4, we present a set of deterministic (without drift, Fig. 4A, B and C) and stochastic (Fig. 4D and 4E) simulations obtained with a random landscape (blue and green dots represent two alternative habitats). We first consider the “instantaneous colonization” model in which simulations start with parental genotypes fixed in each habitat type (Fig. 4A). We believe that this model might illustrate how secondary introgression happens in a patchy environment after successive colonization waves, and how it is often imagined when interpreting the conflictual pattern between neutral and selected loci (Fig. 1). The simulation is deterministic so that the final equilibrium state is genetic homogeneity at the neutral locus. However, spatial and genetic structures substantially delay the homogenization process so that it takes a longer time to reach equilibrium (Barton 1979; Barton et al., 2007). During the early initial phase, habitat patches are introgressed by alleles of the other ecotype and GEAs progressively dissipate although still reflecting the initial specialization of each taxon into its preferred habitat. Introgression is heterogeneous in space and depends on the local structure of the landscape; it is faster in small patches surrounded by a high density of demes of the alternate habitat, which are more impacted by immigration. GEAs vanish in a heterogeneous fashion, and the neutral diversity progressively forms self-organized spatial patterns. Although GEA is observed at a local scale, IBD is observed at a large scale (the shallow parabolic shape in Fig. 4A). Furthermore, the initial association can be lost and therefore reversed GEAs are observed between distant regions. For example, in Fig. 4A, the initial situation (grey) is by construction a strict association between the C allele and habitat 2 (blue dots); however after a few hundreds of generations (blue curves), the association, in addition to being weaker, is reversed in the center of the range (higher frequencies of C in habitat 1, green dots), while it remains in the initial direction in the periphery. This heterogeneity in the direction of GEAs is likely to be

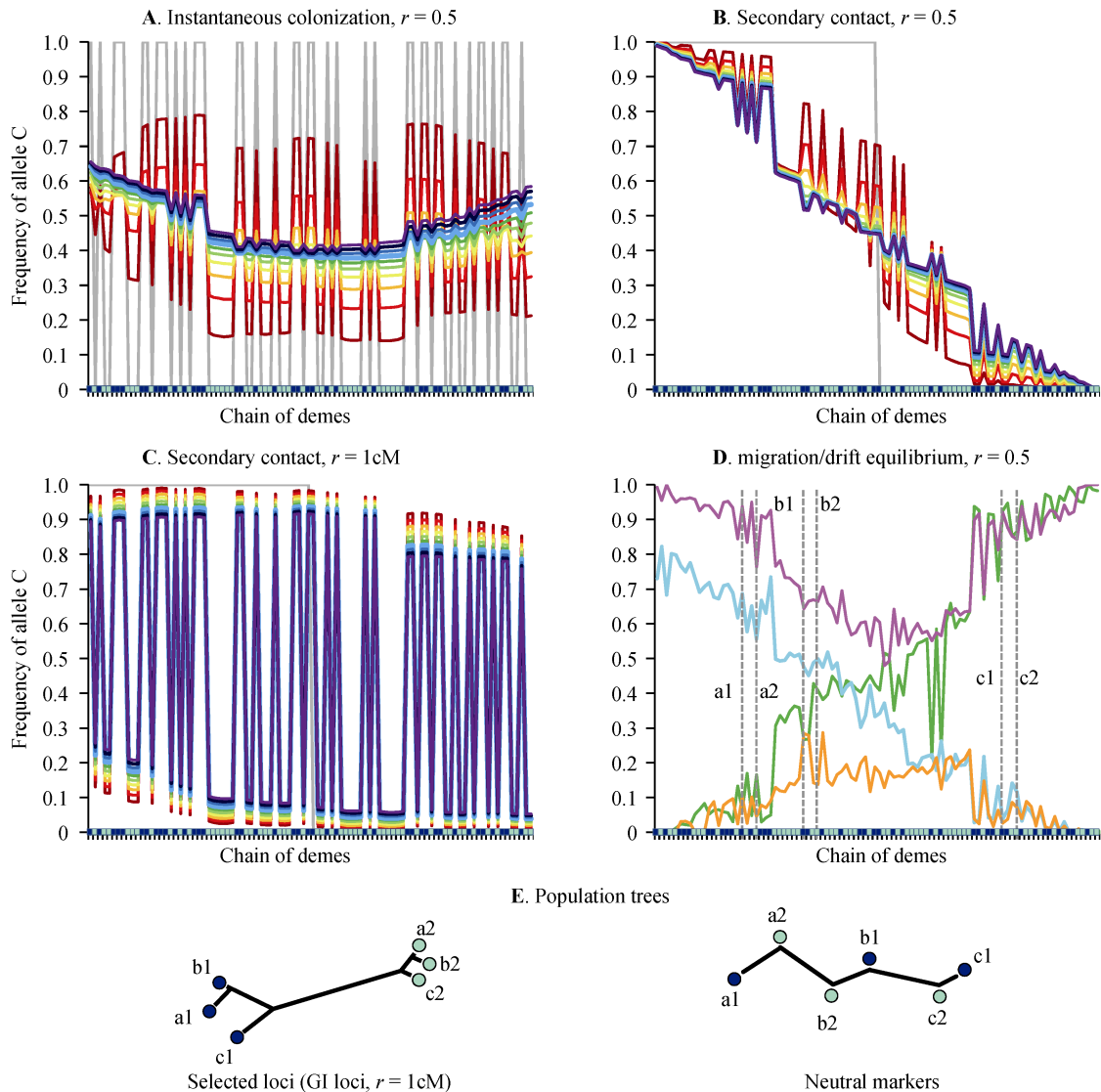


Fig. 4 Simulation outputs obtained with the general model

A. Frequency of the C allele at a neutral locus unlinked to selected loci obtained with the initial conditions of the “instantaneous colonization” model, in which habitat 1 demes (blue) are initially fixed by the ABC genotype and habitat 2 demes (green) with abc. Results plotted every 100 generations from generation 0 (grey) to generation 1000 (purple) following a red to blue rainbow code. **B.** Frequency of the C allele at a neutral locus unlinked to selected loci obtained with the initial conditions of the “secondary contact” model, in which the left part of the chain of demes is initially fixed with the ABC genotype and the right part with abc, independently of the environment. Results plotted every 100 generations from generation 0 (grey) to generation 1000 (purple) following a red to blue rainbow code. **C.** Frequency of the C allele at a neutral locus linked to a selected locus ($r=0.01$) obtained with the initial conditions of the “secondary contact” model. Results plotted every 100 generations from generation 0 (grey) to generation 1000 (purple) following a red to blue rainbow code. **D.** Four independent snapshots of the frequency of the C allele at a neutral locus unlinked to selected loci obtained at migration-selection-drift equilibrium (generation 100 000), the colors of the lines serve to discriminate them and have no meaning. Migration rate $m=0.5$, deme size $N=500$ selection coefficient $s=t=0.9$. **E.** Population trees obtained with the neighbor-joining algorithm on Reynold’s genetic distance obtain with five independent GI loci or five independent neutral markers.

even more erratic under more realistic conditions. Indeed, the realism of the “instantaneous colonization” model can be criticized because the migration rates simulated during introgression would not allow each ecotype to instantaneously colonize its favored habitat everywhere, and therefore in our model the initial state is characterized by artificially clear-cut GEAs.

We now consider a secondary contact model in which

simulations start with one parental genotype fixed in the left part of the chain of demes and the other parental genotype fixed in the right part. In this case, associations between a neutral locus unlinked to a barrier locus and the selected background quickly dissipate during the colonization phase, generating a geographical cline at a large scale (i.e. a special case of IBD pattern) although GEA is observed at a local scale (Fig. 4B). The

neutral locus is affected by the genetic barrier, so that gene flow is locally reduced between habitats, but it is even more affected by historical isolation. For a neutral locus to effectively hitchhike with the selected background during colonization, we would need to simulate a nearly complete barrier, *i.e.* true species (not shown). This illustrates the ease with which recombination breaks the associations between loci and that a genetic barrier is mainly efficient at a local chromosomal scale (Barton, 1979; Charlesworth et al., 1997). Indeed, Fig. 4C shows that linkage allows neutral alleles to hitchhike with selected alleles and to delay the introgression process substantially, generating very strong and persistent GEAs at a global scale, although a weak and large-scale clinal pattern slowly emerges. With varying degrees of linkage between the neutral locus and selected loci, one can expect a range of situations from a predominantly clinal pattern with local and weak GEA (Figure 4B) to a predominantly habitat-driven genetic structure (GEA) with a very weak cline (Fig. 4C). Finally, we introduce random genetic drift and consider the migration-selection-drift equilibrium that can be reached whatever the initial conditions. Fig. 4D shows four independent snapshots (e.g. this could represent four independent loci) of allele frequency variation patterns (for clarity we chose to represent two geographic gradients and two parabolic structures, which are illustrative of the diversity of situations we observed). Again, GEA is observed at a local scale but IBD is observed at a large scale, and reversed GEAs can be observed in different parts of the range (Fig. 4D). To illustrate the conflictual pattern of Fig. 1 we generated population trees by computing neighbor-joining trees on Reynold's genetic distance at five independent neutral and GI loci (as we simulated only one neutral locus, we reiterated the simulations five times) with six populations, three from habitat 1 and three from habitat 2 (see Fig. 4D). Figure 4E shows that based on GI loci (linked to a selected locus with $r = 1\text{cM}$) populations cluster according to habitat type, while the effect of geographic distance is visible within each of the two habitats. On the other hand, differentiation at neutral loci tends to reflect geographic isolation, although populations of alternative habitats are more differentiated than expected by their geographic distance.

2.2 Case study 1: *Littorina saxatilis*

We used the model to investigate how introgression could proceed in the high-shore/low-shore landscape structure faced by *L. saxatilis* ecotypes which is quite similar to our general model although the pair of habi-

tats is uniformly distributed along the distribution range (the coast). In the deterministic model of secondary contact (Fig. 5A) geographical clines are generated by secondary contact (history) and dissipate with time, whereas concomitantly, the genetic barrier locally maintains isolation between habitats. In the stochastic model at equilibrium, the result is similar, geographic isolation maintains IBD at a large scale while GEA is observed at a local scale (Figure 5B). This is also visible in the population trees computed by neighbor-joining on Reynold's genetic distance at five independent neutral and GI loci (Fig. 5C).

Furthermore, the initial association between neutral alleles and the selected background is lost, and associations vary spatially, so that the *C* allele is more frequent in the high shore at some places (peripheral populations in the snapshot of Fig. 5B) and more frequent in the low shore at other places (central populations in the snapshot of Fig. 5B). Examples of reversed GEAs are discussed at length in Bierne et al., (2011). One example that resembles the *L. saxatilis* system is found in the barnacle *Semibalanus balanoides*. In this species, the allozyme locus *Mpi* has been found to be differentiated between high-tide and low-tide microhabitats (Schmidt and Rand, 1999). However, opposite zonation patterns at the *Mpi* locus were observed in Maine (Schmidt and Rand, 1999), Rhode Island (Rand et al., 2002) and the Gulf of St. Lawrence (Véliz et al., 2004).

Coming back to *Littorina*, our simple model of introgression in a patchy environment therefore reproduces the pattern observed in natural populations which has been taken as evidence for ongoing primary differentiation due to ecological speciation (Wilding et al., 2001; Panova et al., 2006; Quesada et al., 2007; Rolán-Alvarez 2007; Galindo et al., 2009; Johannesson et al., 2010). This pattern is expected because geographic isolation (see Glossary) at a large scale is stronger than genetic isolation (see Glossary) at a small scale, and it is obtained whatever the initial conditions.

2.3 Case study 2: Threespine stickleback

The model used to illustrate how introgression could proceed in the marine/freshwater landscape structure faced by sticklebacks is different from previous models. The reason is that direct single generation migration between adjacent streams is made impossible. This creates a strong asymmetry in demographic processes with many small isolated stream populations and a large nearly panmictic marine population, corroborated by the analysis of real populations (e.g. Hohenlohe et al., 2010). Isolation between streams does not prevent the alleles

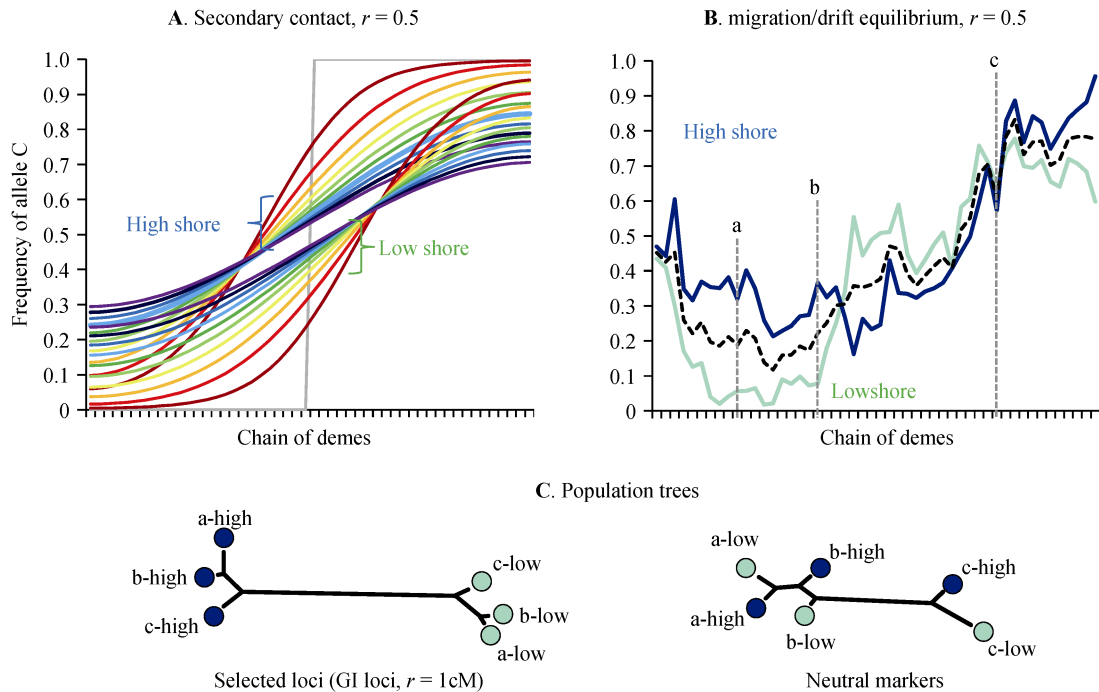


Fig. 5 Simulation outputs obtained with the *Littorina saxatilis* landscape illustrated in Fig. 2A

A. Frequency of the C allele at a neutral locus unlinked to selected loci obtained with the initial conditions of the “secondary contact” model, in which the left part of the chain of demes is initially fixed with the ABC genotype and the right part with abc, independently of the environment. Results plotted every 100 generations from generation 0 (grey) to generation 1000 (purple) following a red to blue rainbow code. **B.** One snapshot of the frequency of the C allele at a neutral locus unlinked to selected loci obtained at migration-selection-drift equilibrium (generation 100 000), blue line in high shore (habitat 1) demes, green line in low shore (habitat 2) demes and dotted black line is the average frequency on the shore. Migration rate $m=0.5$, deme size $n=500$ selection coefficient $s=t=0.9$. **C.** Population trees obtained with the neighbor-joining algorithm on Reynold’s genetic distance obtain with five independent GI loci or five independent neutral markers.

involved in adaptation to freshwater to spread providing selection is not incredibly strong and allows them to be present at low frequency in the marine habitat (Schluter and Conte, 2009). However, it does prevent unlinked neutral loci to efficiently hitchhike with them during the spread (not shown), as it already did in the landscapes models illustrated in Fig. 4B. We can nonetheless assume that the freshwater ecotype could have colonized streams by following meltwater from retreating glaciers after the last ice age, which is the way many freshwater fishes are thought to have colonized northern latitudes (Bernatchez and Wilson, 1998). We therefore consider this initial condition (a freshwater taxon initially fixed in all streams), knowing from previous models that the system will converge to an equilibrium state that is independent of the initial conditions.

Fig. 6 illustrates that introgression (i) proceeds in different streams in a mostly independent fashion, and that freshwater populations also drift independently from one another, and (ii) strongly asymmetrically from the marine to the stream genome. This latter observation, that introgression proceeds asymmetrically from the “big” to the “small” population, is a well-established

theoretical prediction (Barton 1986). At migration-selection-drift equilibrium, the initial association is lost, GEA is observed but with no directionality (the most frequent allele of one stream is not necessarily the same in another stream), and stream genomes have been mostly swamped by marine alleles at neutral loci (Fig. 6). Below each graph is represented the population tree (neighbor-joining on Reynold’s genetic distances) obtained with five independent neutral loci. The historical genetic similarities among stream populations disappear after introgression swamping is finished and the geographic proximity between rivers does not translate into increased genetic similarity.

3 Discussion

Decades of research on “single-cline” hybrid zones have corroborated the theoretical expectation that primary intergradation is difficult to discriminate from and secondary introgression (Barton and Hewitt, 1985). Thorough phylogeographic analyses have nonetheless often confirmed post-glacial secondary contacts (Hewitt, 2000). Moreover, the divergence between parapatric forms often predates the last glacial oscillation so that

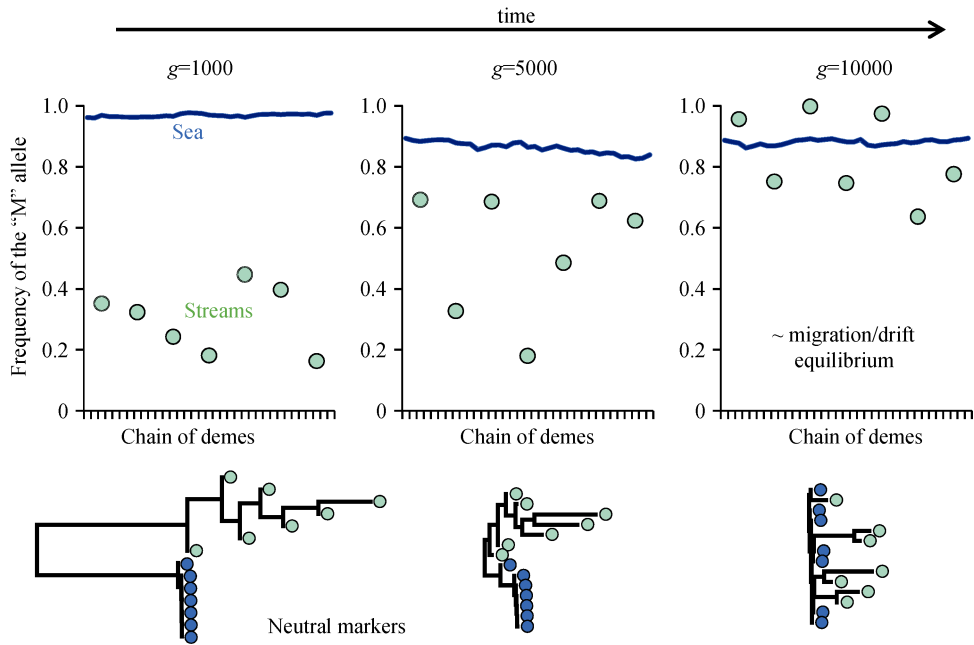


Fig. 6 Simulation outputs obtained with the stickleback landscape illustrated in Fig. 2B

Frequency of the “M” allele at a neutral locus (defined as the neutral allele initially fixed in marine populations) unlinked to selected loci. One snapshot is presented at generation 1,000, 5,000 and 10,000. Blue lines: frequency in marine demes. Green dots: frequency in streams. Below each graph is represented the population tree obtained with five independent neutral loci. Population trees obtained with the neighbor-joining algorithm on Reynold’s genetic distance obtain with five independent neutral markers.

most of the divergence between incipient species has occurred during a succession of contraction/expansion cycles (Hewitt, 2001, 2011). This means that some polymorphisms involved in reproductive isolation and/or adaptation of incipient species to contrasted habitat types have survived for long periods during which populations have changed in size and spatial distribution and that the recent diversity and genetic structure of neutral markers does not necessarily reflect the conditions (allopatry, parapatry or sympatry) that prevailed when they initially emerged. Some authors have therefore suggested that we should abandon the classical parapatric/sympatric dichotomy to integrate divergence on a multidimensional system that accounts for time (history), space (gene flow), evolutionary forces at play (selection *versus* drift), isolation mechanisms (pre-, post-zygotic, endogenous, exogenous) and their genetic architecture (Butlin et al., 2008; Abbott et al., 2013). On the other hand, there is a renewed interest in ecological speciation, with studies focusing on ecologically-driven divergent selection as the primary factor driving the cascade of evolutionary processes leading to speciation (Schluter, 2001; Rundle and Nosil, 2005; Nosil and Feder, 2012). Unfortunately they tend to neglect the complexity of historical processes that determined the current distribution of genetic variants in species complexes.

Empirically, the ecological speciation literature relies on model systems in which reproductive isolation is thought to have evolved recently between ecotypes (e.g. Via, 2009), and does not give full consideration to history and long-term evolutionary processes. One argument used to neglect historical complexity is often based on the conflicting patterns displayed by neutral and selected loci (GEA at selected traits/loci *versus* local GEA at a small spatial scale and IBD at a large spatial scale at neutral loci, Fig. 1). Indeed, it is assumed that in the case of primary differentiation, genes linked to selected polymorphisms involved in the differentiation process should form genomic islands of differentiation driven by local adaptation and forming GEA, while the rest of the genome reflects neutral processes such as isolation by distance. In contrast, in the case of a secondary contact, all neutral genes are expected to follow predominantly concordant clinal patterns of differentiation reflecting the initial position of differentiated allopatric taxa, independently of habitat distribution. In the case of a progressive habitat gradient (e.g. a latitudinal gradient), it can be difficult to distinguish GEA from IBD (e.g. contact zone between a northern and a southern race *versus* local adaptation to temperature from cold northern areas to warm southern areas). However when the distribution of habitats is patchy and fine-grained (e.g. low shore and high shore in *Littorina*)

GEA follows this distribution and cannot be confounded with large-scale IBD, and the contrast between selected and neutral variation is obvious. Here, we argue that this contrast is not a sufficient proof to discriminate primary from secondary intergradations. During secondary intergradation, polymorphisms involved in adaptation to different habitats, that preexisted in one or both ancestral taxa, can come to produce GEA. Neutral loci sufficiently linked to them in genomic islands of differentiation (GI loci) also produce GEA. On the other hand, the genetic structure at neutral loci outside of genomic islands self-organizes independently according to historical vicariance and/or geographic distance. The net effect is the same as parallel primary differentiation. The deep coalescence of adaptive polymorphisms identified so far (Colosimo, 2005; Wood et al., 2008) is rather in accordance with an old divergence of ancient alleles that have survived contraction/expansion cycles.

Most models of genetic barriers have considered a simple distribution of races (two-deme or clinal structure) while few models exist that interpret the genetic structure of partially reproductively isolated taxa that segregate spatially in accordance with a fine-grained mosaic environment. The principal result of our simulation work is that it is difficult to obtain GEA with neutral markers unlinked to a selected locus since this requires very strong selection coefficients (or pre-zygotic isolation) at the genome level. However, when sufficiently strong selection coefficients are used, the genetic barrier still often remains less efficient than geographic isolation. As soon as geographic isolation is stronger than genetic (ecotypic) isolation, a large scale IBD/fine scale GEA pattern is obtained very quickly after a secondary contact and is indefinitely maintained at migration-selection-drift equilibrium. Linkage is very effective in strengthening the genetic barrier in the chromosomal neighborhood of selected loci (as the barrier strength is proportional to the s/r ratio) and, contrary to neutral loci unlinked to selected loci, GI loci (neutral markers belonging to genomic islands of differentiation, see Glossary) retain a large scale GEA pattern for very long periods (Fig. 4C). In other words linkage allows genetic isolation to be stronger than geographic isolation. We therefore conclude that distinguishing primary intergradation from secondary intergradation with the analysis of neutral loci unlinked to selected loci is, contrary to common belief, as difficult in a patchy environment as it is between parapatrically distributed forms.

In *Littorina saxatilis*, contrasting gene/population

trees have been observed based on neutral and selected markers, and although the secondary contact hypothesis was never definitely rejected (Wilding et al., 2001; Grahame et al., 2006; Butlin et al., 2008), this system became a classical example of parallel ecological speciation (Quesada et al., 2007; Rolán-Alvarez, 2007; Galindo et al., 2009; Johannesson et al., 2010). The lack of a larval dispersal phase in this species explains why local adaptation can be maintained against gene flow between high and low-shore populations, but one can also argue that this life history trait might enhance geographic differentiation between distant locations. Our simulations confirm that it is difficult to produce genetic barriers able to overwhelm geographic isolation, and that the contrast between neutral and selected markers is expected under both primary and secondary intergradation, as previously warned (e.g. Grahame et al., 2006). The secondary contact hypothesis could possibly be refuted by the finding of independent genetic bases of local differentiation between habitats across the species range, however parallelism has not yet been tested at the genetic level in this species. This could be done in the near future using candidate genes identified in BACs (Wood et al., 2008) and transcriptome scans (Galindo et al., 2010).

Some speciation genes or genomic regions thought to have evolved as a by-product of adaptation to different environments and to contribute to partial reproductive isolation between incipient species have now been identified (Wolf et al., 2010; Nosil and Schluter, 2011). To date, exceptionally differentiated loci between ecotypes in different habitats have revealed strikingly deep coalescences (Schulte et al., 1997; Pogson, 2001; Colosimo, 2005; Wood et al., 2008), leading to the conclusion that ecological adaptation frequently relies on old polymorphisms, the age of which often predates last quaternary glacial oscillation cycles. As a result, the ecological speciation literature gives increasing consideration to the role of adaptation "from standing genetic variation" (Colosimo, 2005; Barrett and Schluter, 2008a; Schluter and Conte, 2009), although it is not always clear what exactly is meant by this phrase in the context of speciation (see below). Another alternative hypothesis is that speciation was initiated earlier in the past, and that initial isolating barriers have been maintained and subsequently associated with other types of isolating barriers (Barton and de Cara, 2009; Bierne et al., 2011). For example a post-zygotic isolation due to intrinsic genetic incompatibilities, resulting in a relative inability to form viable hybrids can arise in allopatry. These genetic in-

compatibilities can later, when the incipient species come into contact, become associated with habitat-specialization polymorphisms within the contact zone (Bierne et al., 2011).

What are the conceptual differences between the hypothesis of primary differentiation based on selection from standing genetic variation and the hypothesis of a co-existence of two incompatible genetic backgrounds subjected to secondary introgression? Both scenarios rely on the “reuse” of old adaptive polymorphisms and thus they might appear quite similar. However the mind-set is different and we would like to highlight important differences. The best way to illustrate these differences is to take an example. We will use the stickleback example, although the argument is valid for other systems (periwinkles, pea aphids, walking sticks etc.). This example is demonstrative for two reasons. First, the spatial context, with a big panmictic marine population and many small independent stream populations, is at first sight suggestive of primary parallel differentia-

tion. Second, this is the model in which the idea of “reuse of shared standing genetic variation” has been popularized, and this idea has survived the characterization of many barrier loci/regions of deep coalescence. Indeed, despite accumulating evidence for an old age of alleles in the genomic islands of divergence, the standing variation argument has been put forward to maintain the initial hypothesis that adaptation to the stream environment itself was recent (Colosimo et al., 2005; Barrett and Schluter, 2008b; Schluter and Conte, 2009; Jones et al., 2012b). However, we can imagine three different scenarios schematically illustrated in Figure 7. All scenarios share the basic characteristic that alleles conferring adaptation to stream environments are old, as suggested by genomic studies; “old” here means that they predated the establishment of current stream populations which must have occurred after the last glacier retreat. (A) The stream ecotype has survived the last glacial cycle in a stream refuge somewhere in the south and has later colonized northern streams by following the gla-

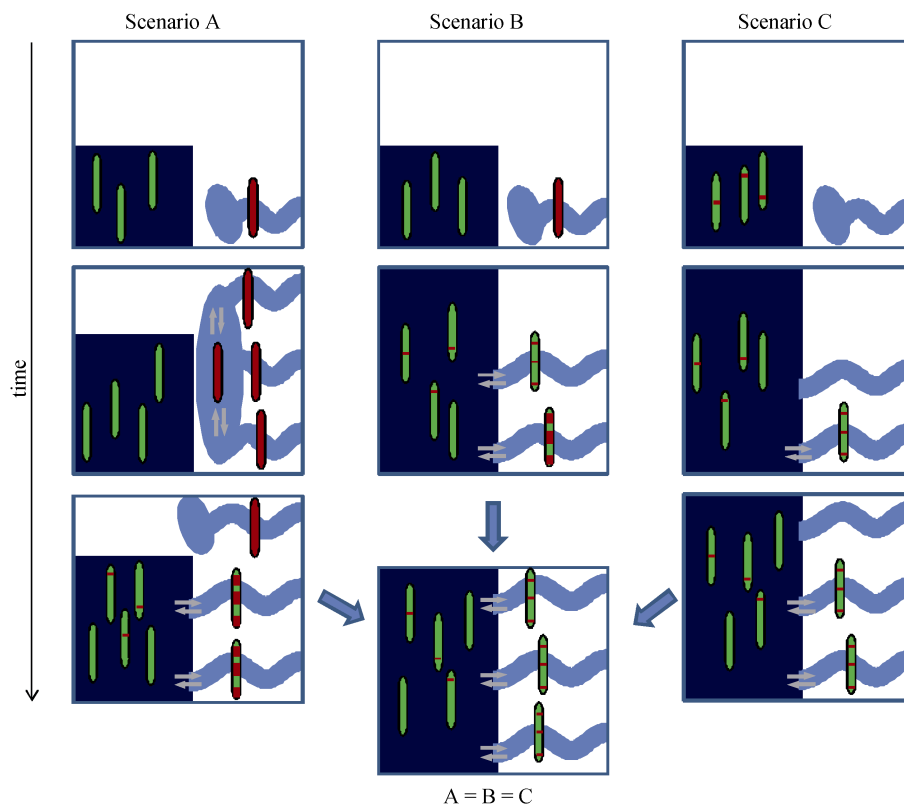


Fig. 7 Simplified schematic illustration of the three scenarios explaining the current genetic structure observed between marine and stream habitats in three-spined sticklebacks

Genomes are depicted by single chromosomes. Only one genome is represented for stream populations (assumed fixed for simplicity) but several genomes are represented for the marine population to illustrate polymorphism. Red: “freshwater” alleles, green: “marine” alleles. Light blue: freshwater environment, dark blue: marine environment. Grey arrows: gene flow. The stream genome survives the last glacial cycle in a refuge in scenarios A and B while stream alleles only survive as standing genetic variation within the marine genome in scenario C. The stream genome as a whole (including neutral alleles) initially colonizes streams in scenario A, while only freshwater-adapted alleles penetrate streams by temporarily introgressing the marine genome in scenario B and C. The three scenarios converge to the same state.

acier meltwater. Under this scenario, stream genomes have subsequently been swamped by marine alleles at genome regions unlinked to selected loci and the remnants of the initial differentiation between the two marine and stream populations are visible only in genomic islands around loci involved in the adaptation to the stream environment, because only these genomic islands resist introgression. This is the situation we modeled in Fig. 5. (B) The stream ecotype has survived the last glacial cycle in a stream refuge somewhere in the south and the freshwater-adapted alleles at selected loci allowed the colonization of northern streams by temporarily introgressing the marine genome, although this introgression did not affect parts of the genome unlinked to selected loci. The stream refuge has in this case been swamped by marine alleles and is now unrecognizable as a refuge, or has not been found yet. (C) Stream populations have totally disappeared during the last glacial cycle and stream alleles have survived as standing genetic variation within the marine genome, allowing the freshwater-adapted genotype to reemerge in newly ice-released streams after the glacier retreat. It is unclear if the hypothesis of selection from standing genetic variation considers scenario C only, or if scenario B also belongs to this hypothesis. The existence of a stream refuge during the last glacial cycle seems however a pivotal question which is independent of the propagation route of the stream genotype (A and B *versus* C). Indeed, only the C scenario can be used as an argument for ongoing ecological speciation; A and B are in effect secondary contacts between anciently diverged genetic backgrounds, which differ only by the point that either most stream populations were once pure and later became introgressed (A) or were introgressed during their propagation from river to river (B). Even if stream alleles are only slightly deleterious in the marine environment and the marine background, it seems difficult to understand why they would have survived for long periods in the marine environment (scenario C); the old age of the alleles revealed by genomic studies so far is expected for protected polymorphisms, not for deleterious alleles. On the other hand, when freshwater-adapted alleles are positively selected in some populations (freshwater refugia) they introgress the marine genome in which they are maintained at low frequency owing to gene flow from freshwater populations, and they can relatively easily spread from river to river passing through seawater populations if they are not under extremely strong counterselection there (scenario B). The existence of a stream refuge is a question that should be

specifically addressed in the stickleback literature. More generally, we believe the hypothesis of parallel adaptation from standing variation diverts the issue from the true question, which is how adaptive polymorphism has been protected during millions of years. Indeed, this explanation is answering an ancillary question: how did or how does the flow of adaptive alleles occur between isolated patches of similar habitat, such as freshwater rivers in the stickleback example. However, adaptation from standing variation does not help to resolve the original question of how and when did the adaptive polymorphisms emerge (was it allopatric, parapatric or sympatric divergence?). Unfortunately, as all scenarios rapidly converge to the same state, we can hardly answer to this question based on present-day patterns of differentiation.

We have until now been mostly concerned by contrasts between patterns at neutral genes and patterns at genes involved in local adaptation (or loci linked to them). However most genetic barriers are multifactorial; there are usually several independent loci involved in adaptation to habitat ("exogenous barrier loci"); and there are also independent loci that create intrinsic genetic pre- or post-zygotic incompatibilities between genomes, not related to habitat ("endogenous barrier loci"). Considering a multilocus and multifactorial genetic barrier, one can ask the two following questions. (i) Are associations between barrier loci (i.e. loci that contribute to isolate two taxa) temporarily broken before and/or during habitat colonization as implicitly assumed in the hypothesis of selection on standing genetic variation, or are they maintained? (ii) Can endogenous pre- and/or post-zygotic incompatibilities (i.e. habitat independent) follow exogenous (i.e. habitat dependent) barrier loci during habitat colonization?

Regarding question (i), we still lack a comprehensive approach to test the ease with which freshwater-adapted alleles that segregate at low frequency in the sea can readily reassemble the freshwater multilocus genotype through migration, recombination and selection. The most recent data in natural populations instead suggest that freshwater alleles remain strongly associated in the oceanic population, as revealed by extensive linkage disequilibrium between them (Hohenlohe, et al., 2012). However, we probably expect linkage disequilibrium in every scenario because of ongoing, if restricted, gene flow between stream and marine populations. As for the spatial structure, linkage disequilibrium converges to an equilibrium value whatever the initial stage. Under primary differentiation the initial state is no association

and linkage disequilibrium builds up. Under secondary contact the initial situation is an elevated linkage disequilibrium which progressively dissipates. Both eventually converge to the same stationary linkage disequilibrium (reflecting migration/selection balance). The stationary value will likely depend, among other things, on the number of generations required to connect adjacent streams. This difficult question could be addressed theoretically, with additional developments since the number of loci involved is likely to be an important parameter. For instance, if stream alleles segregate independently in marine populations, fuelling selection on standing genetic variation, it should be easier to reassociate stream alleles at two loci than at tens of loci. Yet a genome scan conducted between marine and freshwater populations based on whole genome resequencing data has recently revealed a high number of genome regions showing parallelism (Jones et al., 2012b).

It remains uncertain if the parallel differentiation of genomic islands between stream and marine populations results from recent independent selection from standing genetic variation or if two co-adapted gene pools re-organized spatially since the last glaciation and asymmetric introgression has blurred this history at neutral parts of the genome. If co-adapted genetic backgrounds co-exist for long periods while being constantly re-organized in space, it might be misleading to focus on the last spatial re-organization to address the question of the timing of speciation (e.g. to argue it is recent) and the importance of ecology in speciation (e.g. to argue that ecologically based selection is the main factor driving divergence). Via (2009) metaphorically referred to the “spyglass / magnifying glass” distinction, recommending that it was desirable to study the process of ecological speciation through the “magnifying glass” in recently diverged species pairs rather than to observe patterns reflecting ancient differentiation through the “spyglass”. We believe however that even when one thinks being studying speciation with a “magnifying glass” he might unknowingly be looking at anciently diverged genetic backgrounds through a “spyglass”, and this might well be happening in sticklebacks, pea aphids, walking sticks, periwinkles and other model systems of the ecological speciation literature. Speciation genetics remains a retrospective analysis from populations having evolved multiple isolating mechanisms. Frustrating as it might be, the conditions under which isolation polymorphisms have initially evolved are probably no longer accessible to a reconstruction from genetic data analysis.

Question (ii) has been addressed in Bierne et al., (2011), although it has been discussed in a different context. Polymorphisms contributing to genetic isolation can basically be classified as exogenous (subjected to habitat-dependent selection) or endogenous (subjected to habitat-independent selection, i.e. genetic incompatibilities). When exogenous alleles propagate in a fine-grained environmental mosaic, independent endogenous loci (i.e. endogenous loci that are not involved in epistatic interactions with exogenous loci) are left aside and the coupling breaks down. Only locally around the secondary contact zone can efficient coupling be maintained between endogenous and exogenous loci while exogenous alleles invade their preferred patches of habitats in the whole distribution range (Bierne et al., 2011). This creates local GEA at endogenous loci that have been observed in some mosaic hybrid zones (Harrison and Rand, 1989; MacCallum et al., 1998; Bierne et al., 2002). If we model a pair of incompatible alleles at two loci and a local adaptation locus within the stickleback landscape of Fig. 2b, not only very strong selection coefficients but also very strong linkage between endogenous and exogenous loci (resulting in very strong coupling) are required for endogenous alleles to propagate with exogenous alleles in several streams. Otherwise, endogenous loci self-organize according to space independently from the exogenous locus and the environment, and one genotype should ultimately fix if the tension zone is not stabilized geographically at a natural barrier to gene flow or an environmental boundary (Abbott et al., 2013). This result mirrors the expected conditions required for speciation with gene flow - that associations between assortative mating genes and local adaptation loci need to be facilitated by direct selection, linkage, epistasis and/or pleiotropy to persist (Gavrilets, 2004). On the other hand, if history provides initial conditions in which there is linkage disequilibrium between exogenous and endogenous loci, the coupling can persist and strengthen (Kirkpatrick and Ravigné, 2002). Therefore, if independent pre- or post-zygotic endogenous loci contribute to speciation and parallelism is observed at these loci, this should be in accordance with scenario A - stream ecotypes colonized northern streams by following the glacier meltwater, allowing endogenous and exogenous alleles to colonize every stream conjointly. Chromosomal inversions identified by Jones et al. (2012) might well be such intrinsic incompatibilities. We should emphasize however that even with an initial linkage disequilibrium created in allopatry, the coupling between

endogenous and exogenous loci is still often unstable with the landscape of Fig. 3B and it requires strong selection coefficients and weak genetic drift (big stream population sizes) to persist, but conditions exist for which coupling persists without epistasis or physical linkage.

In conclusion, although strongly debated in the past in the analysis of parapatrically distributed forms separated by “single-cline” hybrid zones, the idea that a contrast in the differentiation pattern observed between selected traits / genes and neutral markers can help in discriminating primary from secondary intergradation has been resurrected in the analysis of patchily distributed forms in the ecological speciation literature. We argue that discriminating primary and secondary intergradation on the basis of this contrast alone is as uncertain in a patchy environment as it is between parapatrically distributed forms. Because neutral markers can quickly lose their initial association with the selected backgrounds (Grahame et al., 2006; Roberts et al., 2010), the only solution to progress in the understanding of the history of adaptation/speciation is to reconstruct this history using gene genealogies at the loci under selection or at neutral markers linked to them (i.e. belonging to genomic islands of differentiation). To date, data suggest that these adaptive polymorphisms have had a long history and have survived many spatial re-organizations produced by contraction/expansion cycles in glacial periods, arguing against the hypothesis of recent independent emergence of adaptation to a new habitat type. Furthermore, the number of barrier loci and the complexity of the barrier to gene flow observed in key examples of the ecological speciation literature (e.g. in maggot flies: Michel et al., 2010; in sticklebacks: Jones et al., 2012a), would rather be in accordance with the hypothesis that co-adapted genetic backgrounds have co-existed for long periods. One challenge of ecological speciation will probably be to incorporate history and long-term evolutionary processes into a framework that until now has tended to over-emphasize hypotheses of recent adaptation and fast evolution. One avenue of research could be to investigate the historical divergence and geographical distribution of intrinsically incompatible alleles (endogenous barriers), which probably exist, and investigate if their coupling with local adaptation genes (exogenous barriers) has persisted or has broken down.

Acknowledgements We would like to thank Jenny Boughman and two anonymous reviewers for helpful comments on

the manuscript. This work was funded by the Agence National de la Recherche (Hi-Flo project ANR-08-BLAN-0334-01). This is article 2012-XXX of Institut des Sciences de l'Evolution de Montpellier.

References

- Abbott R, Albach D, Ansell S et al., 2013 Hybridization and speciation. *Journal of Evolutionary Biology in press*.
- Barrett R, Schluter D, 2008a. Adaptation from standing genetic variation. *Trends in Ecology and Evolution* 23: 38–44.
- Barrett RDH, Schluter D, 2008b. Adaptation from standing genetic variation. *Trends in Ecology and Evolution* 23: 38–44.
- Barton NH, 1979. Gene flow past a cline. *Heredity* 43: 333–339.
- Barton NH, 1983. Multilocus clines. *Evolution* 37: 454–471.
- Barton NH, 1986. The effects of linkage and density-dependent regulation on gene flow. *Heredity* 57: 415–426.
- Barton NH, Bengtsson BO, 1986. The barrier to genetic exchange between hybridising populations. *Heredity* 57: 357–376.
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Nipam HP, 2007. *Evolution*. New York: Cold Spring Harbor Laboratory Press.
- Barton NH, de Cara MAR, 2009. The evolution of strong reproductive isolation. *Evolution* 63: 1171–1190.
- Barton NH, Hewitt GM, 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16: 113–148.
- Bernatchez L, Wilson CC, 1998. Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* 7: 431–452.
- Bierne N, David P, Langlade A, Bonhomme F, 2002. Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? *Marine Ecology-Progress Series* 245: 157–170.
- Bierne N, Welch J, Loire E, Bonhomme F, David P, 2011. The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20: 2044–2072.
- Bonin A, Taberlet P, Miaud C, Pompanon F, 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog *Rana temporaria*. *Molecular Biology and Evolution* 23: 773–783.
- Butlin RK, Galindo J, Grahame JW, 2008. Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 2997–3007.
- Campbell D, Bernatchez L, 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution* 21: 945–956.
- Charlesworth B, Nordborg M, Charlesworth D, 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* 70: 155–174.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr., Dickson M et al., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307: 1928–1933.
- Coyne JA, Orr HA, 2004. *Speciation*. Sunderland, MA: Sinauer Associates, Inc..
- Deagle BE, Jones FC, Chan YF, Absher DM, Kingsley DM et al., 2011. Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceed-*

- ings of the Royal Society B: Biological Sciences 279(1732): 1277–1286.
- Egan SP, Nosil P, Funk DJ, 2008. Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* 62: 1162–1181.
- Endler JA, 1977. *Geographic Variation, Speciation and Clines*. Princeton, New Jersey: Princeton University Press.
- Feder JL, Egan SP, Nosil P, 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* 28: 342–350.
- Galindo J, Grahame JW, Butlin RK, 2010. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of Evolutionary Biology* 23: 2004–2016.
- Galindo J, Moran P, Rolan-Alvarez E, 2009. Comparing geographical genetic differentiation between candidate and non-candidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Mol. Ecol.* 18: 919–930.
- Gavrilets S, 2004. *Fitness landscapes and the origin of species*. Princeton and Oxford: Princeton University press.
- Gavrilets S, Cruzan MB, 1998. Neutral gene flow across single locus clines. *Evolution* 52: 1277–1284.
- Grahame JW, Wilding CS, Butlin RK, 2006. Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution* 60: 268–278.
- Harrison RG, 1993. *Hybrid Zones and the Evolutionary Process*. Oxford: Oxford University Press.
- Harrison RG, 2011. Understanding the origin of species: Where have we been, where are we going? In: Bell MA, Futuyma DJ, Eanes WF, Levinton JS ed. *Evolution since Darwin: The First 150 Years*. Sunderland, Massachusetts: Sinauer Ass. Inc., 119–149.
- Harrison RG, 2012. The language of speciation. *Evolution in press*.
- Harrison RG, Rand DM, 1989. Mosaic hybrid zone and the nature of species boundaries. In: Otte D, Endler JA ed. *Speciation and Its Consequences*. Sunderland, Massachusetts: Sinauer Ass. Inc., 111–133.
- Hewitt GM, 2000. The genetic legacy of the quaternary ice ages. *Nature* 405: 907–913.
- Hewitt GM, 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology* 10: 537–549.
- Hewitt GM, 2011. Quaternary phylogeography: The roots of hybrid zones. *Genetica* 139: 617–638.
- Hohenlohe PA, Bassham S, Currey M, Cresko WA, 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 395–408.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA et al., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6, e1000862.
- Johannesson K, Panova M, Kempainen P, André C, Rolán-Alvarez E et al., 2010. Repeated evolution of reproductive isolation in a marine snail: Unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 1735–1747.
- Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD et al., 2012a. A Genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology* 22: 83–90.
- Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E et al., 2012b. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484: 55–61.
- Kirkpatrick M, Ravigné V, 2002. Speciation by natural and sexual selection: Models and experiments. *The American Naturalist* 159: S22–S35.
- Kruuk LEB, Baird SJE, Gale KS, Barton NH, 1999. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* 153: 1959–1971.
- MacCallum CJ, Nurnberger B, Barton NH, Szymura JM, 1998. Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* 52: 227–239.
- Michel AP, Sim S, Powell TH, Taylor MS, Nosil P et al., 2010. Widespread genomic divergence during sympatric speciation. *Proc. Natl. Acad. Sci. USA* 107: 9724–9729.
- Nosil P, Egan SP, Funk DJ, 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation By Adaptation' and multiple roles for divergent selection. *Evolution* 62: 316–336.
- Nosil P, Feder JL, 2012. Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 332–342.
- Nosil P, Schluter D, 2011. The genes underlying the process of speciation. *Trends in Ecology and Evolution* 26: 160–167.
- Panova M, Hollander J, Johannesson K, 2006. Site-specific genetic divergence in parallel hybrid zones suggests nonallopatric evolution of reproductive barriers. *Molecular Ecology* 15: 4021–4031.
- Pogson GH, 2001. Nucleotide polymorphism and natural selection at the Pantophysin, Pan I. Locus in the Atlantic Cod *Gadus morhua* L.. *Genetics* 157: 317–330.
- Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E, 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* 61: 1600–1612.
- Rand DM, Spaeth PS, Sackton TB, Schmidt PS, 2002. Ecological genetics of *Mpi* and *Gpi* polymorphisms in the acorn barnacle and the spatial scale of neutral and non-neutral variation. *Integrative and Comparative Biology* 42: 825–836.
- Roberts DG, Gray CA, West RJ, Ayre DJ, 2010. Marine genetic swamping: hybrids replace an obligately estuarine fish. *Molecular Ecology* 19: 508–520.
- Rolán-Alvarez E, 2007. Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. *Journal of Molluscan Studies* 73: 1–10.
- Rundle HD, Nosil P, 2005. Ecological speciation. *Ecology Letters* 8: 336–352.
- Schluter D, 2001. Ecology and the origin of species. *Trends in Ecology and Evolution* 16: 372–380.
- Schluter D, Conte GL, 2009. *Colloquium Papers: Genetics and ecological speciation*. *Proceedings of the National Academy of*

- Sciences 106: 9955–9962.
- Schmidt PS, Rand DM, 1999. Intertidal microhabitat and selection at *Mpi*: Interlocus contrasts in the northern acorn barnacle *Semibalanus balanoides*. *Evolution* 53: 135–146.
- Schulte PM, Gómez-Chiari M, Powers DA, 1997. Structural and functional differences in the promoter and 5' flanking region of *Ldh-B* within and between populations of the teleost *Fundulus heteroclitus*. *Genetics* 145: 759–769.
- Smadja CM, Butlin RK, 2011. A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology* 20: 5123–5140.
- Véliz D, Bourget E, Bernatchez L, 2004. Regional variation in the spatial scale of selection at *MPI** and *GPI** in the acorn barnacle *Semibalanus balanoides* (Crustacea). *Journal of Evolutionary Biology* 17: 953–966.
- Via S, 2009. Colloquium Papers: Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* 106: 9939–9946.
- Via S, West J, 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* 17: 4334–4345.
- Wilding CS, Butlin RK, Grahame J, 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* 14: 611–619.
- Wolf JBW, Lindell J, Backstrom N, 2010. Speciation genetics: Current status and evolving approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 1717–1733.
- Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK, 2008. Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology* 17: 3123–3135.

GLOSSARY

Barrier loci: Loci at which different alleles are adapted to different environmental conditions or are affecting habitat choice (exogenous barrier loci) and loci that produce hybrid fitness depression or pre-zygotic isolation irrespective of the environment (endogenous barrier loci).

Genetic barrier to gene flow: A reduction in effective gene flow at neutral loci as a consequence of selection on barrier loci.

Neutral loci: Loci that do not affect individual fitness and are not linked to a barrier locus. Note that in the context of our study, such loci are still indirectly affected by the genome-wide genetic barrier to gene flow imposed by barrier loci.

Genomic islands of differentiation: Portions of chromosome affected by a barrier locus. Theory predicts that a selected locus (locally adapted or intrinsically incompatible) should decrease the effective migration rate of linked neutral loci proportionally to the ratio s/r (Barton, 1979b; Charlesworth et al., 1997). Because of the inverse relation with r , islands of differentiation are expected to be small (no more than 20 cM), which can result in large physical islands in regions of low recombination rates. Alternatively, a succession of linked barrier loci on the chromosome can also produce large islands (Via, 2012).

Genomic island loci (GI loci): Loci that do not affect individual fitness but belong to a genomic island of differentiation – i.e. that are linked to a barrier locus. Note that in our simulations, GI loci loosely linked to a barrier locus have an intermediate behaviour between neutral and barrier loci, being equally affected by geographic and genetic isolations, but the neutral/GI dichotomy approximation simplifies the explanation.

Geographic isolation: Genetic differentiation produced by the geographic distance separating populations.

Genetic isolation: Genetic differentiation produced by genetic barriers to gene flow.

Fine-grained environment: Spatial heterogeneity in habitat that occurs at a fine spatial scale relative to species dispersal, so that migration often occurs between contrasting habitats.

Co-adapted genomes or backgrounds: An assemblage of alleles at barrier loci that evolved together and have been selected to work together (optimised fitness).

Parallel adaptation: The independent evolution of the same or similar multilocus genotypes at barrier loci due to the similarity of selection regimes and/or habitats in distant locations.

Standing variation: The segregation of alleles at a locus within a population; in the context of ecological speciation, this phrase often refers to the presence of alleles adapted to another habitat, segregating at low frequency in a population in which it does not confer local adaptation.