

INVITED REVIEW

The coupling hypothesis: why genome scans may fail to map local adaptation genes

NICOLAS BIERNE,*† JOHN WELCH,*†§ ETIENNE LOIRE,†‡ FRANÇOIS BONHOMME*† and PATRICE DAVID*¶

*Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France, †ISEM – CNRS, UMR5554, SMEL, 2 rue des Chantiers, 34200 Sète, France, ‡AD2M – CNRS, UMR7144, Station Biologique, 29682 Roscoff Cedex, France, §Department of Genetics, University of Cambridge, Cambridge CB23EH, UK, ¶CEFE – CNRS, 34293 Montpellier Cedex 5, France

Abstract

Genomic scans of multiple populations often reveal marker loci with greatly increased differentiation between populations. Often this differentiation coincides in space with contrasts in ecological factors, forming a genetic–environment association (GEA). GEAs imply a role for local adaptation, and so it is tempting to conclude that the strongly differentiated markers are themselves under ecologically based divergent selection, or are closely linked to loci under such selection. Here, we highlight an alternative and neglected explanation: intrinsic (i.e. environment-independent) pre- or post-zygotic genetic incompatibilities rather than local adaptation can be responsible for increased differentiation. Intrinsic genetic incompatibilities create endogenous barriers to gene flow, also known as tension zones, whose location can shift over time. However, tension zones have a tendency to become trapped by, and therefore to coincide with, exogenous barriers due to ecological selection. This coupling of endogenous and exogenous barriers can occur easily in spatially subdivided populations, even if the loci involved are unlinked. The result is that local adaptation explains where genetic breaks are positioned, but not necessarily their existence, which can be best explained by endogenous incompatibilities. More precisely, we show that (i) the coupling of endogenous and exogenous barriers can easily occur even when ecological selection is weak; (ii) when environmental heterogeneity is fine-grained, GEAs can emerge at incompatibility loci, but only locally, in places where habitats and gene pools are sufficiently intermingled to maintain linkage disequilibria between genetic incompatibilities, local-adaptation genes and neutral loci. Furthermore, the association between the locally adapted and intrinsically incompatible alleles (i.e. the sign of linkage disequilibrium between endogenous and exogenous loci) is arbitrary and can form in either direction. Reviewing results from the literature, we find that many predictions of our model are supported, including endogenous genetic barriers that coincide with environmental boundaries, local GEA in mosaic hybrid zones, and inverted or modified GEAs at distant locations. We argue that endogenous genetic barriers are often more likely than local adaptation to explain the majority of F_{st} -outlying loci observed in genome scan approaches – even when these are correlated to environmental variables.

Keywords: ecological speciation, endogenous barrier, F_{st} outliers, genetic incompatibilities, genetic–environment association, genome scan, hybrid zone, local adaptation

Received 11 June 2010; revision received 14 February 2011; accepted 26 February 2011

Introduction

Evolutionary processes are most often inferred indirectly from contemporary patterns of genetic variation (Harrison 1993, 1998). Here, we consider a pattern that has long attracted the attention of population geneticists: the association between the level of genetic differentiation at specific loci (e.g. those identified as outliers by *F_{st}* scans) and one or more ecological variables (genetic–environment association, GEA, Hedrick *et al.* 1976). GEA is a pattern that is well explained by the existence of ecologically driven divergent selection at some genes in the genome, which in itself warrants interest. However, the processes by which the signal of local selection can be captured by *F_{st}*-outlying markers may not be as straightforward as is commonly believed.

The simplest hypothesis is that the loci detected are directly involved in local adaptation. Examples now exist where the whole chain of links between genotype, phenotype, fitness and environmental variation has been characterized (Lenormand *et al.* 1999; Hoekstra *et al.* 2004; Wheat *et al.* 2006; Schmidt *et al.* 2008b; Storz & Wheat 2010), but they remain rare. Often genetic markers are neutral and reveal local adaptation through their association with selected loci. The reliance on this indirect path (i.e. the use of neutral markers to infer the behaviour of selected loci) remains widespread, despite our ability in the age of genomics to scan hundreds of loci with ease. Unfortunately the indirect nature of the inference is often neglected (Faure *et al.* 2008; Bierne 2010), and the indirect effect of selection on linked neutral variation is implicitly assumed to be similar in kind (if not strength) to that of direct selection. For example it is not uncommon to read that the very strong genetic structure observed at a given locus is consistent with the operation of selection on that locus *or a locus closely linked to it*, as if the direct or indirect nature of selection was an ancillary detail. However, the indirect effect of selection on neutral markers requires a different theory than direct selection. It is better understood as a stochastic process, which can be approximated as within-genome variation in the effective population size (Felsenstein 1974; Gillespie 2000; Charlesworth 2009) or in the effective migration rate between populations (Barton 1979b; Bengtsson 1985; Ingvarsson & Whitlock 2000). Moreover, the ability of selection to affect population differentiation at neutral markers is not yet fully understood. For example, the hitchhiking effect, in which the fixation of a favourable gene produces a marked and durable footprint on linked neutral variation (Maynard & Haigh 1974; Nielsen 2005), can also have a strong effect on genetic differentiation between demes in a structured population, and these are less well studied (Slatkin & Wiehe 1998; Barton 2000; Santiago & Cabal-

lero 2005; Bierne 2010). The effect of local selection on the linked neutral variation is expected to extend over a very small chromosomal region (Charlesworth *et al.* 1997; Bierne 2010; Feder & Nosil 2010) and to last for a short period of time (Miller & Hawthorne 2005) – predictions supported by the few empirical examples in which the chromosomal footprint of local adaptation has been investigated (Berry *et al.* 1991; Schmidt *et al.* 2008b; Linnen *et al.* 2009). Recent extensive genome scans have shown that genetic differentiation does not extend far beyond 5 kb (<0.003% of the genome) around adaptive polymorphisms in *Arabidopsis lyrata* (Turner *et al.* 2010) and *Drosophila melanogaster* (Kolaczowski *et al.* 2010). With this expectation in mind, and given the number of loci usually studied (almost always a small fraction of the genome), invoking some form of association between the marker and a linked selected locus is often no more parsimonious than assuming direct selection on the marker itself (Lemaire *et al.* 2000). Indeed, direct selection tended to be the favoured hypothesis in the golden age of allozyme loci (Koehn & Hilbish 1987; Johannesson *et al.* 1995a; Mitton 1997; Lemaire *et al.* 2000; Riginos *et al.* 2002), and remains very popular as an interpretation of results obtained with EST-derived markers (markers in protein coding sequences, Vasemägi *et al.* 2005; Oetjen & Reusch 2007; Galindo *et al.* 2010; Shikano *et al.* 2010). However, loci with an abnormally strong genetic structure, the so-called *F_{st}* outlier loci, are common among both coding and non-coding polymorphisms, and often represent a substantial fraction of the loci investigated (2–10%, Nosil *et al.* 2009). If each locus only maps 0.003% of the genome as in the rock-cress and fly genomes, thousands of polymorphic local-adaptation loci would have to be present to account for the observed proportions of outliers. We therefore suggest that the standard explanation, as outlined above, cannot plausibly explain why GEAs are found so easily and commonly with molecular markers.

Various alternative hypotheses have already been proposed to explain outlying loci: (i) correlation in co-ancestry among subpopulations under some forms of spatial subdivision (e.g. historical branching, hierarchical structure) can inflate the neutral variance of genetic differentiation, which then exceeds that of the null model used in outlier tests (Robertson 1975; Excoffier *et al.* 2009; Bonhomme *et al.* 2010); (ii) a neutral mutation that arises in the wave front of an expanding population can increase in frequency in colonized populations, mimicking local selection at that locus (gene surfing, Klopstein *et al.* 2006; Hofer *et al.* 2009); (iii) for a broad range of parameter values, background selection against deleterious mutations is expected to increase population differentiation because of decreased

effective size within subpopulations (Charlesworth *et al.* 1997), sometimes to decrease it because of heterotic effects (Bierne *et al.* 2002c), and in any case to inflate the differentiation variance when loci from regions with different recombination rates are analysed; (iv) transient differentiation can arise from species-wide selective sweeps (Bierne 2010), or multiple sweeps (Ralph & Coop 2010), by globally favourable mutations (not locally adapted alleles) in subdivided species. In this review we advocate another alternative and neglected explanation, that a high proportion of F_{st} outliers likely imply the operation of an endogenous genetic barrier. We argue this because (i) endogenous barriers are more likely to impede neutral gene flow at a substantial proportion of the genome (Barton & Bengtsson 1986); (ii) they are expected to become coupled with exogenous selection, and thus to coincide with environmental variation (Barton & Hewitt 1985), forming GEAs.

Throughout, incompatibilities between groups of alleles due to habitat-independent selection (underdominance or epistasis) or pre-zygotic isolation will be referred to as *endogenous barriers to gene flow*, and where populations with such incompatible genetic backgrounds come into contact they form a *tension zone*. Groups of alleles adapted to different habitats form *exogenous barriers to gene flow*, and their geographical limits are habitat boundaries or *ecotones*. Endogenous and exogenous genetic barriers effectively restrict gene flow because of the indirect effect of selection on the linked neutral variation. Genetic barriers are often semi-permeable (Harrison 1986; Rieseberg *et al.* 1999), and their *strength* varies among genome regions depending on local recombination rates and densities of selected loci (Barton & Hewitt 1985). A third kind of barrier to gene flow is natural obstacles (i.e. zones of low population density such as mountains, rivers and oceanic front), which may or may not coincide with ecotones. Unlike a genetic barrier, a natural barrier decreases gene flow equally throughout the whole genome. The terminology used in this paper is defined in the Glossary.

Tension zones, being due to endogenous selection, are not stabilized geographically and, in the absence of exogenous selection, they are expected to move more or less haphazardly according to population density, stabilizing only when they reach natural barriers, with which they will ultimately coincide (Barton 1979a; Barton & Hewitt 1985; Hewitt 1988). However, speciation and hybrid zone theories have highlighted that different components of reproductive isolation can easily become coupled together in the absence of natural barriers (Udovic 1980; Kirkpatrick & Ravigné 2001; Barton & de Cara 2009). The most discussed type of coupling is between post-zygotic (e.g. hybrid fitness depression or disruptive selection) and pre-zygotic (e.g. assortative mating) factors, because

this underlies the reinforcement (Butlin 1989) and ecological (Rundle & Nosil 2005) models of speciation. The coupling between multiple endogenous genetic incompatibilities has also long been recognized in the hybrid zone literature (Barton 1983; Barton & Hewitt 1985), and Barton & de Cara (2009) recently emphasized its role in speciation. The coupling between endogenous and exogenous components of reproductive isolation is a further prediction of the theory, but this has not tended to be strongly emphasized to date. For example, the concomitant action of endogenous and exogenous factors has often been discussed in the hybrid zone literature (Barton & Hewitt 1985; Hewitt 1988; Moore & Price 1993), but the discussions have focused on the relative importance of the two factors (Arnold 1997; Barton 2001) rather than on their possible interaction. In addition, the concept of genetic barriers has not sufficiently permeated the literature on local adaptation at the molecular level, being primarily restricted to the hybrid zone literature. However, hybrid zones are little more than extreme examples of the broader phenomenon of genetic differentiation. Genome scans simply reveal more subtle examples: weak barriers that affect a smaller but still substantial proportion of genomes (>2%). In the latter case, we suggest that the exclusive focus on local adaptation as the cause of the barrier might be mistaken.

In this article, we present a simple model in which endogenous and exogenous selection interact. We emphasize two underappreciated outcomes of this interaction: (i) endogenous clines (i.e. clines of allele frequencies at genes under endogenous selection) can come to coincide with exogenous clines (i.e. clines of allele frequencies at genes under exogenous selection) at an environmental boundary; (ii) fine-grained GEAs can emerge locally within an endogenous cline. We then review results from the literature that support our hypothesis: (i) theory suggests that endogenous barriers are more efficient than exogenous barriers at impeding neutral gene flow in a substantial proportion of the genome, (ii) genetic barriers to gene flow are usually both exogenous and endogenous, (iii) endogenous barriers tend to have environmental/physical correlates, (iv) GEAs are often evident only at local scales in mosaic hybrid zones, and (v) the association between genetic differentiation and habitat variation is sometimes reversed or modified at distant locations.

The barrier trap: coupling exogenous and endogenous selection

In this section we present a simple model to illustrate the coupling of endogenous and exogenous backgrounds, and how this coupling shapes the spatial genetic structure.

A simple model

We present a model of evolution in a metapopulation of n demes arranged in a linear stepping-stone structure. Migration connects not only adjacent demes but also demes separated by up to six demes. 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.

We consider three unlinked bi-allelic haploid loci. The first two loci have alleles labelled A and a (locus 1) and B and b (locus 2). These two loci are involved in a symmetric genetic incompatibility (endogenous selection) in which a is incompatible with B , and b is incompatible with A . The two-locus fitnesses are $W(AB) = W(ab) = 1$ and $W(Ab) = W(aB) = 1 - s$, in all demes in the chain (here we modelled strong endogenous selection, $s > 0.5$). Below, we model a secondary contact between these two endogenous backgrounds (AB and ab) by initially fixing AB in the left part of the chain of demes and ab in the right part of the chain.

The third locus has alleles labelled C and c , which are under exogenous selection. Each deme in the chain was assigned to one of two habitats types (habitat 1 and habitat 2), and one of the alleles was adapted to each habitat. Allelic fitness were $W(C) = 1$ and $W(c) = 1 - t$ in habitat 1 and $W(C) = 1 - t$ and $W(c) = 1$ in habitat 2 (t was set to 0.1 for scenarios with moderate exogenous selection and to 0.9 for scenarios with strong selection). The assignment of habitat types, and the initial state of locus 3 varies in the simulations presented below.

The effects of endogenous and exogenous selection combined multiplicatively so that, for example, $W(ABC) = (1 - s)(1 - t)$ in habitat 1. Genotypic frequencies in each deme were derived from those of the previous generation after accounting for recombination, migration and selection in that order. Simulations were often deterministic (i.e. deme sizes were assumed to be large enough for drift to be negligible) but random drift was also simulated by multinomial sampling of genotypes within each deme at each generation ($N = 200$ per deme). Windows executables are provided as supplementary material and Borland Delphi 4.0 source code is available from the authors upon request.

Moderate selection: tension zones are trapped by exogenous clines

Selective coupling arises in the form of linkage disequilibrium among barrier loci when increased variance in

compatibility increases mean fitness (multiplicative fitness or positive epistasis, Barton & de Cara 2009). In a spatial model, barrier loci organize in clines maintained by the *dispersal/selection balance*. Linkage disequilibria are caused by dispersal even with no linkage, which favours the coupling process. Coupling therefore happens when clines overlap (Slatkin 1975). For our purposes, the problem is to explain why and how endogenous and exogenous clines can overlap. We here propose three different scenarios in which endogenous clines are produced by a secondary contact of incompatible backgrounds. However, multilocus endogenous clines can also accumulate in parapatry and this will be described elsewhere.

In the first scenario, environmental variation is fine-grained, and each deme is assigned to habitat 1 or habitat 2 at random. We then model a secondary contact between two backgrounds (ABC and abc) differentially adapted to this environmental variation. Figure 1 shows the results of two simulations under this scenario. While the endogenous loci form a cline at the position of the contact (thin line in Fig. 1), the exogenous alleles (C and c) quickly spread through the metapopulation because they are favoured in one of the local habitat types, found at either side of the cline. In the initial stages of the spread, because selection is moderate ($s = 0.5$, $t = 0.1$), the initial association between endogenous and exogenous alleles is lost, and spatial structures establish themselves quite independently. However, once the spatial structure has been established (and to a lesser extent during its establishment) linkage disequilibrium between endogenous and exogenous loci builds up and spatial coupling occurs. As the position of the endogenous cline is not constrained by selection, it can move, and eventually comes to rest where the gradient in exogenous allele frequency is locally maximal (bold lines in Fig. 1). As expected, the tension zone moves so as to minimize its length (Barton & Hewitt 1985). Interestingly, the eventual phase of the coupling (whether AB becomes associated with C or with c) does not depend strongly on the initial phase. Here, all simulations began with only ABC and abc genotypes, and yet many ended with ABc and abC being most common. Simulations which led to both outcomes are shown in Fig. 1a and b.

Figure 2 depicts a second scenario: secondary contact between two backgrounds adapted to habitat 2 (genotypes Abc and abc) in a two-patch coarse-grained environment. All demes on the left-hand-side of the chain are habitat 1, and those on the right, habitat 2. We assume that allele C (which is favoured in habitat 1) now arises in the first deme on the left side (in the habitat 1 patch) at an initial frequency $p_0 = 0.001$. The spread of this allele throughout the habitat 1 demes is

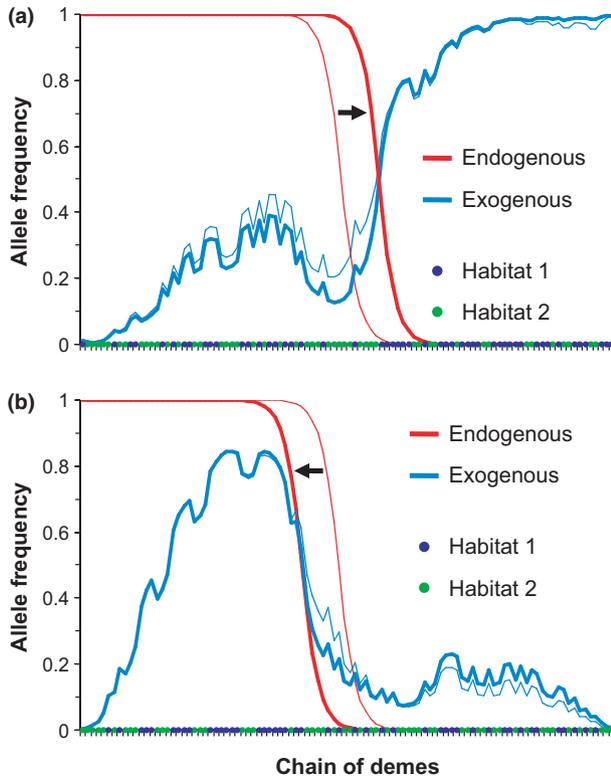


Fig. 1 Secondary contact between two intrinsically incompatible backgrounds in the presence of fine-grained environmental variation. Each background is initially adapted to one of two alternative habitats, and each deme is assigned to habitat 1 (blue dots) or habitat 2 (green dots) at random. Selection is moderate ($s = 0.5$, $t = 0.1$). Migration connects demes separated by up to six demes (see text). Exogenous alleles (blue lines) quickly spread through the metapopulation and organize according to the landscape (thin blue line). The endogenous loci form a cline at the position of the contact (thin red line), then coupling occurs and the endogenous cline can move and comes to rest where the gradient in exogenous allele frequency is locally maximal (bold red line). Once coupling is complete, the spatial variation of the exogenous alleles is slightly modified by the barrier to gene flow generated by the endogenous cline (bold blue line). The black arrow illustrates the movement of the endogenous cline. (a) and (b) show individual simulations in which the endogenous background from the demes on the left coupled with the exogenous allele favoured in habitat 2 (a) or habitat 1 (b).

virtually unaffected by the endogenous barrier, as is expected for a favourable allele (Barton 1979b; Pialek & Barton 1997), but the endogenous cline moves slightly as allele C crosses because the AB background becomes transiently associated with C. The outcome then depends on the distance of the environmental boundary from the endogenous cline. If this distance is too great, then the coupling does not persist and the wave of advance continues toward the environmental boundary leaving the endogenous cline behind (not shown).

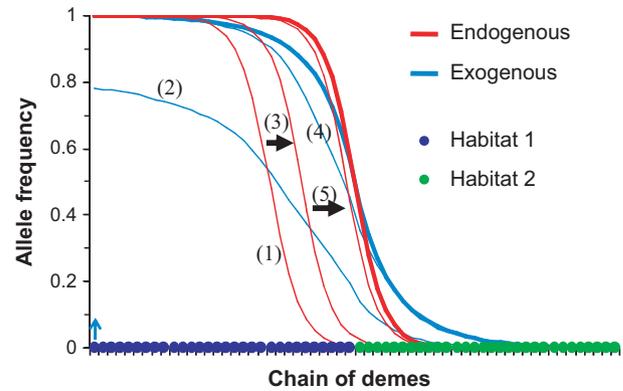


Fig. 2 Effect of the spread of a new locally favoured mutation on a secondary contact endogenous cline in a two-patch coarse-grained environment. A chain of 60 demes was divided into 30-deme patches of habitat 1 (blue dots) or habitat 2 (green dots). Selection is moderate ($s = 0.5$, $t = 0.05$). Migration connects demes separated by up to six demes (see text). The endogenous cline (red line) is initially positioned at deme 20 within the habitat 1 patch (1). A new exogenous allele arises in the first deme on the left (blue arrow) and spreads throughout habitat 1, crossing the endogenous cline (2), which transiently moves (3), and forms a cline at the environmental boundary (4). As the endogenous and exogenous clines overlap, coupling continues and the endogenous cline moves until it coincides with the environmental boundary (5). The width of the exogenous cline is reduced due to coupling with the endogenous cline. Hence, the exogenous and endogenous clines coincide with each other and with the environmental boundary (bold lines). Black arrows illustrate the movement of the endogenous cline.

Figure 2 shows a situation where the coupling does persist, that is, when the exogenous cline, once established at the environmental boundary, overlaps with the endogenous cline so that some demes are polymorphic for both endogenous and exogenous loci (Slatkin 1975; Barton & de Cara 2009). In this case, the endogenous cline moves until it coincides with the environmental boundary where the gradient in exogenous allele frequency is maximal and the width of the endogenous cline is minimal (Barton & Hewitt 1985). In this scenario the phase of the association is always the same (AB with C) owing to the initial position of AB populations (initially predominant in habitat 1, where C is favoured). An important characteristic of this process is that coupling occurs because the exogenous and endogenous clines overlap in space. Although stronger exogenous selection favours the coupling process, it also reduces the width of the exogenous cline, and thus opportunities of overlap. Paradoxically, coupling is therefore more likely with moderate exogenous selection as it results in very wide cline (e.g. latitudinal clines) able to trap the endogenous tension zone from a remote distance.

Genetic drift does not modify the outcomes of the two previous examples qualitatively. With drift the efficacy of selection is reduced and therefore the coupling process is weakened. However drift can also help the endogenous cline to move haphazardly in space, which increases its probability of reaching the attraction basin of exogenous clines near ecotones. Our third scenario, depicted in Fig. 3, illustrates this point. We consider an endogenous cline that moves because of stochastic processes. We modelled an environmental pocket of habitat 1 demes surrounded by two patches of habitat 2 demes, implying that the endogenous cline could be trapped by one of the two environmental boundaries without reaching the ends of the range. The initial position and width of exogenous and endogenous clines does not allow coupling to occur without some movement of the endogenous cline (they do not overlap). The endogenous cline moves randomly because of drift while the positions of exogenous clines are stabilized by selection. The endogenous cline is certain eventually to overlap with one of the two exogenous clines, allowing coupling to occur, and the two clines to collapse at one of the environmental boundaries. Figure 3 shows the position of the endogenous cline at secondary contact (thin line) and of the exogenous and endogenous clines after coupling (bold lines). The process is best observed by using the Windows executable provided as a supplementary file.

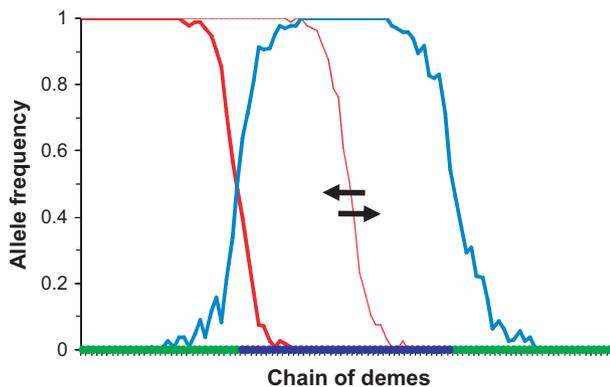


Fig. 3 Barrier trap. A secondary contact results in an endogenous cline at the centre of a chain of 100 demes (thin red line), in between two exogenous clines (blue line). Selection is moderate ($s = 0.5$, $t = 0.1$). Genetic drift occurs in demes of $N = 200$ individuals. Migration connects demes separated by up to six demes (see text). The endogenous cline moves randomly because of drift (illustrated by the two black arrows) while the positions of exogenous clines are stabilized by selection. The endogenous cline eventually comes to overlap with one of the two exogenous clines, allowing coupling to occur, and the two clines collapse at the environmental boundary (bold lines).

Strong selection in a fine-grained environment: GEAs at endogenous loci within a cline

We now model strong endogenous selection. We assume large selection coefficients on a small number of loci for convenience, but the same selection intensity could have been modelled with many loci, each with small effects. In this latter case each locus is expected to cumulate indirect selective effects from other loci in addition to its own selection coefficient (Barton 1983; Kruuk *et al.* 1999a) resulting in so-called 'congealed genomes' (Turner 1967) that behave roughly as a single superlocus under strong selection (Kruuk *et al.* 1999a).

First, as in Fig. 1, we consider secondary contact between two backgrounds (*ABC* and *abc*) differentially adapted to a fine-grained environment. In this case, even strong endogenous selection cannot prevent favourable exogenous alleles from rapidly crossing the tension zone and spreading throughout the metapopulation (Barton 1979b). However, within the tension zone a small-scale GEA arises, producing a mosaic structure in which allele frequencies at endogenous loci locally co-vary with habitat type (as in Fig. 4a). Contrary to the outcome with moderate selection (Fig. 1), the phase of the initial association (*AB* associated with *C*) is nearly always retained. Another interesting result is that GEA at endogenous loci can only be observed locally within the tension zone because this is the only place where these loci are polymorphic, although both habitat heterogeneity and polymorphism at exogenous loci (i.e. local adaptation) are present elsewhere in the range (Fig. 4). If neutral gene flow is prevented by the endogenous barrier, GEA will also be observed locally (in the tension zone) with neutral markers. We here have concentrated on the case where exogenous and endogenous loci do not interact directly but one can imagine the situation in which exogenous alleles are adapted to their native endogenous background while being deleterious in the foreign background (i.e. epistasis). In this case, and if epistasis is sufficiently strong, exogenous alleles can remain confined in their native background, being unable to cross the tension zone (not shown, Bierne 2001).

In the above scenario, each endogenous background is initially adapted to a different habitat. As the environmental heterogeneity predates the secondary contact, and populations of each background live in both habitats prior to the contact, this assumption is questionable. Locally adaptive polymorphism is not unlikely to be present in both endogenous backgrounds before they come into contact. In a second category of simulations, we consider a model in which habitat polymorphism at the exogenous locus occurs within the two endogenous backgrounds. This simulates the situation in which

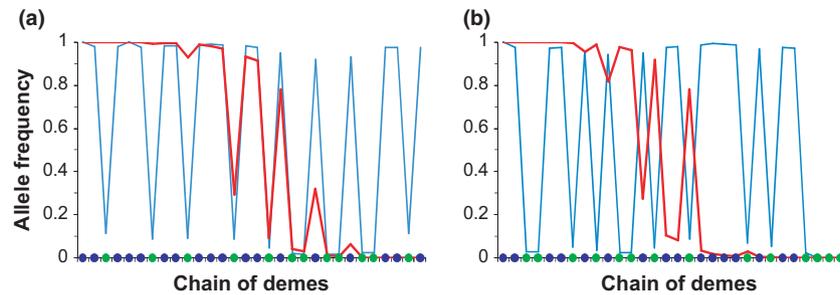


Fig. 4 Secondary contact between two intrinsically incompatible backgrounds when polymorphism at the exogenous locus segregated within the two endogenous backgrounds. Each deme is assigned to habitat 1 (blue dots) or habitat 2 (green dots) at random. Selection is strong ($s = 0.9$, $t = 0.9$). Migration connects demes separated by up to six demes (see text). Exogenous alleles rapidly spread throughout the metapopulation (blue lines), crossing the tension zone like favourable alleles. Coupling occurs locally within the tension zone and a mosaic structure emerges at the endogenous loci (red lines). (a) and (b) show simulations in which the endogenous background from the left becomes coupled with the exogenous allele favoured in habitat 1 (a) or habitat 2 (b).

adaptation to micro-environmental heterogeneity predated the secondary contact. In this model, small-scale GEA readily emerged in the tension zone as previously, but the phase of the coupling (whether AB associates with C or c) proved to form in any direction depending on the environmental landscape and initial configurations. Simulations leading to opposite phases of association are shown in Fig. 4a and b.

Relevant theory and data from the literature

In this section we review theory and data that support the hypothesis of coupling between endogenous genetic barriers and environmental variation. Although many of these arguments have previously appeared in the hybrid zone literature (Barton & Hewitt 1985), they are also valid for more weakly isolated backgrounds that would not usually be called hybrid zones.

Endogenous genetic barriers are probably more efficient than local adaptation at preventing neutral gene flow at a substantial portion of the genome

Anybody who has modelled a barrier to gene flow imposed by a selected locus on a linked neutral locus, whether via underdominance or disruptive selection, must have been impressed, maybe disappointed, by the facility with which recombination breaks the association between the two loci, and how small the chromosomal portion affected by selection is. Indeed, the strength of a genetic barrier is expected to be roughly proportional to $1/r$ (Barton 1979b) and therefore to quickly decrease with the recombination rate.

Barton & Bengtsson (1986) identified the conditions required for the flow of neutral genes to be significantly reduced. First, the barrier needs to be produced by many loci. Thus, one can define two multilocus geno-

types or sets of alleles (genetic backgrounds) that, upon interbreeding, produce a variety of intermediate genotypes called hybrids. Second, the combination of alleles from the two backgrounds in hybrid genotypes must substantially reduce the fitness (hybrid fitness depression). This does not imply that selection must be endogenous, only that hybrids should perform poorly whatever the environment (while each parental type outperforms the other in a given habitat). Note that hybrid unfitness is a strong prerequisite of models of ecological speciation (Gavrilets & Vose 2005); in this sense, ecological speciation is nothing other than a form of reinforcement in which hybrid unfitness is caused by exogenous rather than endogenous selection. Hybrids represent 'bridges' between genetic backgrounds and they need to be rare for a barrier to be effective. Selection against hybrids also generates a reduction in population density that mechanically impedes gene flow, the 'hybrid sink' effect (Barton 1980, 1986).

Barton & Bengtsson (1986) emphasized that for the barrier to be genome-wide 'the number of genes involved in building the barrier must be so large that the majority of other genes become closely linked to some locus which is under selection'. If only a fraction of loci screened exhibit high levels of differentiation, the requirement is that the number of genes involved in the barrier must be large enough for these markers to be closely linked to a selected locus. Recall, however, that F_{st} outliers often represent an appreciable proportion of the panel of loci screened (2–10%).

Theory thus predicts that hybrid fitness depression distributed across many loci is the most efficient selection regime to prevent the flow of neutral genes. Which of endogenous and exogenous selection is the most likely candidate for such effects? We argue that endogenous selection on intrinsic genetic incompatibilities is more plausible than ecologically driven divergent

selection. It is now well established that hybrid fitness depression between well defined taxa results from the combination in hybrid genotypes of alleles at two or more loci involved in negative epistatic interactions, so-called Dobzhansky–Muller (DM) incompatibilities (Orr & Presgraves 2000; Coyne & Orr 2004; Gavrillets 2004). The accumulation of alleles involved in DM incompatibilities does not require environmental changes and can occur whenever populations happen to be geographically isolated (Orr 1995) including in parapatry (Gavrillets *et al.* 2000; Kondrashov 2003; Navarro & Barton 2003). DM incompatibilities can affect any interacting loci (Noor & Feder 2006; Barton & de Cara 2009) which implies that virtually any mutation in a functional region might become involved in an endogenous barrier. The study of the genetics of post-zygotic reproductive barriers has indeed revealed unforeseen molecular mechanisms (e.g. gene transposition, heterochromatin formation, splicing regulation) underlying the few genetic incompatibilities identified thus far (Masly *et al.* 2006; Ferree & Barbash 2009; Chou *et al.* 2010), which suggests we still have many other fascinating mechanisms to uncover. The DM model therefore provides a simple answer to the existence of efficient genetic barriers to gene flow not only between well delineated species or subspecies (Coyne & Orr 2004), but also between more weakly isolated genetic backgrounds. Conversely, the proportion of sites in a genome that are functionally involved in local adaptation, and can directly lead to an F_{st} outlier, is probably low (Remold & Lenski 2001). To make this claim is not to deny the importance of ecology-driven selection at the individual level, nor to underestimate the absolute number of potential targets of exogenous selection in a genome. Undoubtedly local adaptation can involve changes in many complex phenotypic traits (morphological, life history, behavioural and physiological) each influenced by variation in numerous genes (e.g. Bernatchez *et al.* 2010). But if quantitative trait variation involves many genes, local adaptation is likely to proceed through small allele frequency differences at many loci, and so will not result in a large F_{st} at any locus (Le Corre & Kremer 2003). In addition, we would argue that the genome-wide mutational opportunities for intrinsic incompatibilities are probably even more numerous than mutational opportunities for local adaptation to the sort considered in most F_{st} scan studies. Genomes are immensely large, most of the genes or ‘elements’ (transposons, chromosomal arrangements, duplications, splicing sites, satellite repeats, recombination hotspots, retroviruses, vertically transmitted symbionts, etc.) are modestly affected by the extrinsic environment while they are involved in complex epistatic and pleiotropic, co-adaptive or antagonistic

(arms-race) interactions with other genes or elements – in other words they adapt to their evolving intrinsic environment. The likelihood of a molecular marker being closely linked to a gene or element driven by genomic conflicts (Rice & Holland 1997) and involved in a genetic incompatibility (Phadnis & Orr 2009) is probably as high, at least, as the probability of its being closely linked to a gene involved in adaptation to temperature or salinity, notwithstanding that environmental variation is often faced through phenotypic plasticity.

We also acknowledge that the limit between endogenous and exogenous selection is not clear-cut and that we have in this review idealized the distinction for clarity. Alleles adapted to one environment can also be favoured only within their initial genetic background, so that they contribute to endogenous as well as exogenous selection. Furthermore, ecological and endogenous selections are not necessarily independent. Barton (2001) has argued that endogenous isolation is likely to evolve when selection favours different optima for a given trait in different environments, while other traits have the same optimum everywhere. This occurs when mutations have pleiotropic effects on several traits, a common situation, as suggested by measures of genetic correlations in quantitative genetic studies. While populations adapt to different environments by fixing new alleles at loci controlling the trait under divergent selection, compensatory fixations occur at other loci so as to maintain the other traits near their optimum. These compensatory fixations become DM incompatibilities, as they are favoured in the genetic background of the population in which they evolved but potentially detrimental in other backgrounds, irrespectively of the environment. Under Barton’s model, one might expect differential adaptation to generate a genetic barrier composed of many more endogenous than exogenous loci.

Intensively studied genetic barriers are usually multifactorial, and both exogenous and endogenous

The hybrid zone literature has long debated the relative roles of endogenous and exogenous selection (Arnold 1997; Barton 2001). In their seminal review, Barton & Hewitt (1985) concluded: ‘It is harder to distinguish whether parapatrically distributed forms remain distinct because they are adapted to different environments, or because hybrids between them are less fit. However, both direct evidence of hybrid unfitness and the indirect evidence of the close concordance of different characters lead us to believe that the latter is more likely, and that most hybrid zones are in fact tension zones’. However, pronounced GEAs have been identified in

many if not most hybrid zones (Hewitt 1988), at both large (Hunt & Selander 1973; Moore & Price 1993; Riginos & Cunningham 2005; Leaché & Cole 2007) and smaller spatial scales (Rand & Harrison 1989; Howard & Waring 1991; Bierne *et al.* 2002b; Vines *et al.* 2003). Hybrid zones in which the genetic structure is intricately associated with a patchy fine-grained environment have been called mosaic hybrid zones (Harrison & Rand 1989) and undoubtedly imply the action of exogenous factors.

In many cases both exogenous and endogenous factors are acting concomitantly. This observation is far from novel (Barton & Hewitt 1985; Hewitt 1988) but often receives insufficient attention. A list of examples is provided in Table 1. Examples that will be described in more detail below could also be added to this list. The number of multifactorial genetic barriers is remarkable when one considers that the study of hybrid fitness depression is difficult and that neither an absence of fitness depression of intermediate genotypes at marker

Table 1 Examples of multifactorial genetic barriers to gene flow

Hybrid zone	Endogenous factor(s)	Exogenous factor(s)
<i>Bombina bombina</i> / <i>B. variegata</i> (fire-bellied toads)	Selection against hybrids ¹	Habitat preference ²
<i>Mus m. musculus</i> / <i>M. m. domesticus</i> (house mice)	Assortative mating ³ Hybrid infertility ⁴	Association with rainfall ⁵
<i>Littorina saxatilis</i> (rough periwinkle)	Assortative mating ⁶ Selection against hybrids ⁷	Local adaptation ⁸ Habitat preference ⁹
<i>Gryllus firmus</i> / <i>G. Pennsylvanicus</i> (field crickets)	Reproductive incompatibility ¹⁰ Temporal isolation ¹¹	Association with soil type ¹²
<i>Mytilus edulis</i> / <i>M. galloprovincialis</i> (Smooth-shelled blue mussels)	Asynchronous spawning ¹³ Assortative fertilization ¹⁴ Selection against hybrids ¹⁵	Habitat preference ¹⁶ Local adaptation ¹⁷
<i>Mytilus edulis</i> / <i>M. trossulus</i> (Smooth-shelled blue mussels)	Gamete incompatibility ¹⁸ Selection against hybrids ¹⁹	Association with salinity ²⁰
<i>Ostrinia nubilalis</i> (maize/mugwort races of borers)	Assortative mating ^{21,22} Temporal isolation ²¹	Local adaptation ²³
<i>Anopheles gambiae</i> (M and S forms)	Assortative mating ²⁴	Local adaptation ^{25,26}
<i>Colaptes auratus</i> (red/yellow shafted flickers)	Selection against hybrids ²⁷ Territorial defence ²⁸	Association with a multifactorial ecotone ²⁸
<i>Coregonus clupeaformis</i> (normal/dwarf whitefishes)	Selection against hybrids ²⁹	Association with bathymetry ³⁰ Extrinsic hybrid unfitnes ³¹
<i>Salvelinus fontinalis</i> (anadromous/resident brook charrs)	Selection against hybrids ³²	Anadromy/residency tactic ³³
<i>Caledia captive</i> (Moreton/Torresian chromosomal races of grasshopper)	Selection against hybrids ³⁴	Association with a climatic gradient ³⁵
<i>Melanoplus sanguinipes</i> / <i>M. devastator</i> (acridid grasshopper)	Selection against hybrids ³⁶ Assortative mating ³⁶	Association with an altitudinal gradient ³⁶
<i>Dichroplus pratensis</i> (melanopline grasshopper)	Selection against hybrids ³⁷	Association with altitude ³⁸
<i>Heliconius erato</i> / <i>H. melpomene</i> (butterflies)	Selection against hybrids ^{39,40}	Adaptation to local mimicked species ^{39,40}
<i>Acyrtosiphon pisum</i> (Alfalfa/Red Clover races of pea aphids)	Selection against hybrids ⁴¹	Habitat preference ⁴² Local adaptation ⁴¹
<i>Timema cristinae</i> (walking-sticks)	Assortative mating ⁴³	Habitat preference ⁴⁴ Local adaptation ⁴³
<i>Drosophila yakuba</i> / <i>D. santomea</i>	Assortative mating ⁴⁵ Selection against hybrids ⁴⁵	Adaptation to temperature ⁴⁶
<i>Mercenaria mercenaria</i> / <i>M. campechiensis</i> (clams)	Selection against hybrids ⁴⁷	Local adaptation ⁴⁷

¹Kruuk *et al.* (1999b); ²MacCallum *et al.* (1998); ³Smadja *et al.* (2004); ⁴Britton-Davidian *et al.* (2005); ⁵Hunt & Selander (1973); ⁶Pickles & Grahame (1999); ⁷Hull *et al.* (1996); ⁸Janson & Sundberg (1983); ⁹Cruz *et al.* (2004a); ¹⁰Harrison (1983); ¹¹Harrison (1985); ¹²Rand & Harrison (1989); ¹³Gardner & Skibinski (1990); ¹⁴Bierne *et al.* (2002a); ¹⁵Bierne *et al.* (2006); ¹⁶Bierne *et al.* (2003b); ¹⁷Gilg & Hilbish (2003); ¹⁸Rawson *et al.* (2003); ¹⁹Miranda *et al.* (2010); ²⁰Riginos & Cunningham (2005); ²¹Dopman *et al.* (2010); ²²Malausa *et al.* (2005); ²³Calcagno *et al.* (2001); ²⁴White *et al.* (2010); ²⁵Diabaté *et al.* (2005); ²⁶Simard *et al.* (2009); ²⁷Moore & Koenig (1986); ²⁸Moore & Price (1993); ²⁹Lu & Bernatchez (1998); ³⁰Pigeon *et al.* (1997); ³¹Rogers & Bernatchez (2006); ³²Mavarez *et al.* (2009); ³³Thériault *et al.* (2007); ³⁴Shaw & Wilkinson (1980); ³⁵Shaw *et al.* (1993); ³⁶Orr (1996); ³⁷Bidau (1990); ³⁸Tosto & Bidau (1991); ³⁹Mallet (1989); ⁴⁰Mallet & Barton (1989); ⁴¹Via *et al.* (2000); ⁴²Via (1999); ⁴³Nosil *et al.* (2002); ⁴⁴Nosil *et al.* (2006); ⁴⁵Matute & Coyne (2010); ⁴⁶Matute *et al.* (2009); ⁴⁷Bert & Arnold (1995).

loci inferred in nature, nor an absence of hybrid fitness depression in the F1 generation, allows us to definitively rule out endogenous selection (Bierne *et al.* 2006). The observation that tension zones are expected to be trapped by exogenous clines, reconciles Barton and Hewitt's famous assertion – that most genetic barriers are in fact tension zones – with the widespread existence of GEAs: although exogenous factors sometimes explain the location of genetic clines or shifts, they are not necessarily responsible for the barrier to gene flow, and thus for abrupt shifts in allele frequencies at neutral loci. The barrier is more likely due to endogenous selection.

Genetic barriers tend to have physical/environmental correlates

The next question is whether endogenous barriers tend to coincide with environmental limits. There exists an abundant literature on the existence of clusters of hybrid zones or hotspots of genetic structure, sometimes called suture zones (Remington 1968; Hewitt 1996, 2000, 2004; Avise 2000; Swenson & Howard 2005). Questions about suture zones concern their existence (Swenson & Howard 2004), their origins (primary or secondary contacts, Endler 1977) and their maintenance (physical, environmental or genetic barriers). We will here focus on explanations of their position. Three hypotheses have been proposed (i) suture zones occur on zones of secondary contact between glacial refugia, (ii) they represent zones of reduced dispersal (i.e. natural barriers to gene flow) that have trapped multiple tension zones, (iii) they represent environmental boundaries that favour differential adaptation/speciation. We here add a fourth hypothesis: (iv) they represent environmental boundaries that have trapped tension zones through the coupling between endogenous and exogenous backgrounds.

It is clear that suture zones are often well explained as common zones of secondary contact between glacial refugia (Hewitt 2000). For instance, the geography of Europe with its three Mediterranean peninsulas (Iberian, Italian and Greek) that each could have served as refugia during glacial maxima, has most likely resulted in secondary contacts in Southern France and in Central Europe. Similarly in the sea, one can easily imagine that numerous species have been split between the Mediterranean Sea and the Atlantic coasts of Africa during glacial maxima and that populations have secondarily met somewhere near Gibraltar (Patarnello *et al.* 2007). However, this hypothesis cannot explain why genetic structure has not vanished since the secondary contact, why the zones coincide so neatly at a small spatial scale among numerous species and why their position corre-

lates so well with natural barriers to dispersal (e.g. the Alps or the Almeria-Oran front), that are often environmental boundaries as well.

Studying the dynamics of tension zones, Barton (1979a) provided a crucial theoretical prediction: tension zones can move in response to variation in population density and dispersal rate but they are expected to be trapped by natural barriers to dispersal. This prediction not only explains why clusters of hybrid zones often coincide with natural barriers, but it also explains why species with a similar biology (e.g. similar dispersal abilities) can be affected differently by the same barrier. Often, the natural barrier itself would not reduce gene flow enough to generate a strong genetic differentiation, and genetic differentiation can be maintained only because a genetic barrier is superimposed on the natural barrier. This process has been well recognized to explain terrestrial hotspots of hybrid zones (Barton & Hewitt 1985; Swenson & Howard 2005). The Alps in Europe and the Appalachian and Rocky Mountains in USA are such hotspots of hybrid zones well explained by Barton's prediction. Surprisingly, Barton's trapping hypothesis has been relatively neglected in the marine literature (Avise 1992; Palumbi 2003; Patarnello *et al.* 2007; Schmidt *et al.* 2008a) while good examples also exist in the sea. The Almeria-Oran front, an oceanographic front that separates Atlantic from Mediterranean water masses, is a hotspot of genetic structure in the Sea (Patarnello *et al.* 2007). The Almeria-Oran front has been recognized as a natural barrier to dispersal for marine species that disperse via a planctonic larval stage. It is tempting to infer that the front itself is responsible for the genetic structure. However, examples exist of species that do not exhibit any genetic break at the Almeria-Oran front (Launey *et al.* 2002; Patarnello *et al.* 2007) while their dispersal capabilities and population sizes are similar to those of species that do exhibit a break. It is therefore very likely that we presently observe only the tension zones that have been trapped by the front. The same reasoning holds for other hydrographic barriers such as Cape Canaveral in Florida (Avise 1992; Cunningham & Collins 1994), the Siculo-Tunisian Strait in the Mediterranean Sea (Bahri-Sfar *et al.* 2000), or point Conception in California (Burton 1998). Barton's trapping hypothesis can also explain why the locations of shifts in allele frequency do not always coincide between groups of species (e.g. Pelc *et al.* 2009). Species with different dispersal capabilities will not respond identically to natural barriers, and so tension zones of restricted dispersers are expected to be trapped by the first minor barrier encountered near a zone of secondary contact, while tension zones of high dispersers are expected to be

trapped only by a strong natural barrier, even if this is located far from the secondary contact.

While Barton's prediction accounts for the position of many hybrid zones and hotspots of hybrid zones, in other cases the positions of hybrid zones or genetic breaks correspond to ecotone rather than natural obstacles (Moore & Price 1993; Johannesson & Andre 2006), and exogenous selection is assumed to cause genetic differentiation. For instance, Moore & Price (1993) studied in great details the ecological correlates, both abiotic and biotic, of a flicker (Aves; Piciformes) hybrid zone. This zone is localized in the Great Plains of the USA, where there is no obvious natural barrier to dispersal but an obvious ecotone and biogeographic boundary. This area coincides roughly with one of Remington's (1968) suture zones, and has been identified as a hotspot of avian hybrid zones by Moore & Price (1993), as confirmed by the meta-analysis of Swenson & Howard (2005). In this case, it seems clear that ecological selection plays a role in determining the position of the zone. It is also tempting to infer that exogenous selection explains the genetic structure. However, endogenous factors have also been identified, such as low hybrid male fecundity (Moore & Koenig 1986) and assortative mating (Wiebe 2000). An alternative explanation is therefore that the flicker hybrid zone and other hybrid zones in this area are tension zones that have been trapped by exogenous loci at an environmental boundary.

Another example of a hotspot of genetic differentiation with a strong environmental correlate is the Öresund and Danish Belts between the Kattegat (North Sea) and the Baltic Sea (Johannesson & Andre 2006). The most spectacular ecological difference between the Baltic Sea and the North Sea is in salinity. The Baltic Sea is brackish with salinity below 10‰ (as low as 2–4‰ in the northern part) while the North Sea has a standard salinity (~30‰) and the Kattegat is intermediate (~20‰). The Kattegat and the Baltic Sea are separated by a steep salinity gradient. There are also other gradients such as in water temperature, oxygenation and biotic factors, but salinity is usually strongly emphasized. As a consequence, local adaptation to a marginal environment has been proposed as the causative agent of the genetic differentiation, often restricted to a few outlying markers, observed between populations of fish and invertebrates between the Baltic and North Seas (Johannesson *et al.* 1990; Riginos & Cunningham 2005; Johannesson & Andre 2006; Hemmer-Hansen *et al.* 2007b; Gaggiotti *et al.* 2009; Limborg *et al.* 2009). In addition, the Baltic Sea used to be a freshwater lake during the last glacial period and the connection to the North Sea was established only about 8000 years ago, allowing colonization by marine taxa. This has been

taken as an argument against scenarios involving secondary contacts. However this argument neglects the possibility that secondary contact zones can move if they are tension zones and therefore could have formed elsewhere and come to coincide with the environmental boundary secondarily. In fact, many of the genetic deviants found in the Baltic Sea are also observed elsewhere, often in the Barents Sea (northern Scandinavia). Baltic *Mytilus* mussels are *M. trossulus* and form a hybrid zone with *M. edulis* in the Öresund (Väinölä & Hvilsum 1991). *M. trossulus* is also found in some Norwegian fjords (Ridgway & Nævdal 2004; Väinölä & Strelkov 2011), in the Barents Sea (Bufalova *et al.* 2005; Väinölä & Strelkov 2011) and sometimes in the White Sea (Daguin 2000; Väinölä & Strelkov 2011) in Europe as well as on the Pacific and Atlantic sides of North America (Riginos & Cunningham 2005). The tellinid bivalve *Macoma balthica* also forms a hybrid zone in the Öresund (Nikula *et al.* 2008), which is 'replicated' in the Barents Sea such that populations of the Baltic Sea are very similar to populations of the White Sea (Nikula *et al.* 2007; Strelkov *et al.* 2007). Cod (*Gadus morhua*) of the Baltic Sea and of the Barents Sea have similarly low levels of HbI-1 allele at the Haemoglobin-I locus which strongly differentiates cods of the Baltic and North Sea (Petersen & Steffensen 2003). European flounders (*Platichthys flesus*) of the Baltic Sea and of the Barents Sea have similar Hsc70 allele frequencies while both populations are strongly differentiated from populations of the North Sea at this locus (Hemmer-Hansen *et al.* 2007a). In the Amphipods *Gammarus zaddachi*, a cline in allele frequency is observed at the arginine phosphokinase (APK) allozyme locus in the Öresund, yet a population from Tromsø in Norway was found to have the same allele frequencies at the APK locus as populations of the Baltic Sea (Bulnheim & Scholl 1981). Although some might see the footprint of parallel adaptation in such a pattern, the most parsimonious explanation is that a shared history of secondary contact has resulted in multiple tension zones that have all been trapped in the entrance of the Baltic Sea, in the Öresund and Danish Belts. The geography in this region might have promoted a split of a tension zone while it was moving northward, producing two 'daughter' zones, one that became stuck at the entrance of the Baltic Sea and the other going further north along the Norwegian coast. The trapping process could have been produced either by exogenous clines at the ecotone or by a natural barrier to dispersal. Indeed, although the environmental differences between the Baltic and North Sea have been strongly emphasized, the entrance of the Baltic Sea also acts as a physical barrier to dispersal. The steep salinity gradient exists because the water masses do not greatly mix. Either scenario implies that clines now observed in

allele frequencies at a few marker loci in this region do not necessarily reflect salinity-dependent selection at these or linked loci, although adaptation to salinity probably exists somewhere in the genome.

To summarize, barriers to gene flow often have either or both physical and environmental correlates. When they coincide with a natural barrier to dispersal, it is tempting to attribute the genetic structure to the natural barrier. However, it is often more likely that the genetic barrier is a pre-existing tension zone that has become trapped by the natural barrier (Barton's trapping prediction). The natural barrier explains the position of the genetic break but not its maintenance. Similarly, when a barrier to gene flow coincides with an environmental boundary, it is tempting to attribute the genetic structure to local adaptation. However, environmental boundaries often coincide with barriers to dispersal, and even when this is not the case, tension zones can still be trapped by the environmental boundary through the coupling of endogenous and exogenous clines. Local adaptation explains the position of the genetic break but not necessarily its maintenance.

GEAs are evident at a local scale in mosaic hybrid zones

GEAs sometimes involve microhabitat differentiation at a small spatial scale (Table 2). A puzzling observation of such fine-scaled mosaic hybrid zones is that microhabitat heterogeneity (which explains the genetic patchiness within the zone) often exists outside of the zone

as well, but there, does not coincide with genetic heterogeneity at neutral marker loci.

In the large mosaic hybrid zone (spanning from the Southwest of France to the North of Great Britain) between the marine mussels *Mytilus edulis* and *M. galloprovincialis*, habitat specialization is evident between sheltered habitats under freshwater influence, which are occupied by *M. edulis*-like genotypes, and oceanic habitats exposed to wave action, which are occupied by *M. galloprovincialis*-like genotypes (Gardner 1994; Bierne *et al.* 2002b). In the *Gryllus* mosaic hybrid zone, *G. pennsylvanicus* alleles are more frequent in loam soils, and *G. firmus* alleles more frequent in sand soils (Rand & Harrison 1989). In the *Bombina* hybrid zone, *B. bombina*-like individuals are more frequent in ponds, and *B. variegata*-like more frequent in puddles (MacCallum *et al.* 1998; Vines *et al.* 2003). The intuitive interpretation is that *M. edulis* is adapted to sheltered habitats and *M. galloprovincialis* adapted to exposed habitats; that *G. pennsylvanicus* is adapted to loam soils and *G. firmus* to sand soils; and that *B. bombina* is adapted to ponds and *B. variegata* to puddles. However, the sheltered/exposed rocky shores seascape is widespread along the European coastlines, and sand/loam and pond/puddles are also widespread landscapes. The same is true of many of the microhabitat differentiation examples listed in Table 2. The model presented in the previous section provides an alternative interpretation to simple differential adaptation and can also explain why GEAs are observed only locally, in the mosaic hybrid zone, even if the microhabitat heterogeneity is present

Table 2 Examples of microhabitat differentiation

Hybrid zone	Microhabitat differentiation	References
<i>Bombina bombina</i> / <i>B. variegata</i> (fire-bellied toads)	Ponds/puddles	Vines <i>et al.</i> (2003)
<i>Gryllus firmus</i> / <i>G. Pennsylvanicus</i> (field crickets)	Sands/loams	Rand & Harrison (1989)
<i>Mytilus edulis</i> / <i>M. galloprovincialis</i> (Smooth-shelled blue mussels)	High tide/low tide Exposed/sheltered Low salinity/high salinity	Gardner (1994)
<i>Mytilus edulis</i> / <i>M. trossulus</i> (Smooth-shelled blue mussels)	Low salinity/high salinity	Riginos & Cunningham (2005)
<i>Ostrinia nubilalis</i> (European corn borer)	Maize/mugwort	Malausla <i>et al.</i> (2005)
<i>Acyrtosiphon pisum</i> (Pea Aphids)	Alfalfa/Red Clover	Via (2009)
<i>Rhagoletis pomonella</i> (Apple maggot)	Hawthorn/apple	Michel <i>et al.</i> (2010)
<i>Allonemobius fasciatus</i> / <i>A. socius</i> (Ground crickets)	Microclimatic, topographic diversity	Howard & Waring (1991)
<i>Festuca ovina</i> (Sheep's Fescue)	Grassland habitat variation	Prentice <i>et al.</i> (1995)
<i>Semibalanus balanoides</i> (Northern acorn barnacles)	Exposed to the sun, high-tide (thermally stressed)/algal cover, shadow, low-tide	Schmidt & Rand (2001)
<i>Littorina saxatilis</i> (Rough periwinkle)	High shore/low shore	Butlin <i>et al.</i> (2008b)
<i>Fucus spiralis</i> (spiral wrack)	High shore/mid shore	Billard <i>et al.</i> (2010)
<i>Gadus morhua</i> (Atlantic cod)	Coastal/pelagic Depth	Sarvas & Fevolden (2005)
<i>Dicentrarchus labrax</i> (Sea bass)	Sea/lagoon	Lemaire <i>et al.</i> (2000)

elsewhere. Subspecies may be equally adapted to both habitats because habitat adaptation polymorphisms segregate in both backgrounds. However linkage disequilibria between endogenous, exogenous and neutral loci are maintained locally in the hybrid zone, allowing neutral molecular markers to capture the GEA only there.

A posteriori this explanation makes sense for two reasons. First, in their allopatric ranges, each species/subspecies tends to occupy all habitats, with specialization becoming apparent only in contact zones. Polymorphisms for adaptation to different habitats can remain invisible at the level of neutral markers within allopatric populations of either background because no linkage disequilibrium exists between neutral genes and habitat-adaptation genes. Second, even with a very low hybridization rate, the adaptive introgression of exogenous alleles from one background to the other poses no problem as alleles are beneficial in one of the two habitats and it is well known that favourable alleles easily cross tension zones (Pialek & Barton 1997).

The remarkable geographic structure of the hybrid zone between *Mytilus edulis* and *M. galloprovincialis* has largely inspired the ideas developed in this paper. Geographically, this zone is a two-scale mosaic (Bierne *et al.* 2002b). At a large-scale along the Atlantic coast of France, we observed roughly three independent hybrid zones that define two enclosed patches of parental populations in Brittany (*M. galloprovincialis*) and in the Bay of Biscay (*M. edulis*) separated from their external conspecific populations by an allospecific patch (Bierne *et al.* 2003c). Environmental factors do not easily explain this pattern. At a fine-scale, within each hybrid zone, hybrid populations and pure populations of both species are found in close proximity, forming a micro-mosaic structure that correlates with habitat heterogeneity as explained above. The *Bombina* hybrid zone also has the structure of a two-scale mosaic although the large scale mosaic has been attributed to altitudinal differences (Hofman *et al.* 2007). We have argued that the interaction between endogenous and exogenous factors can well explain the fine-grained mosaic structure within a single mosaic hybrid zone, but why are small-scale mosaics repeated independently at several locations and not restricted to a single area? We propose that stochasticity in migration/colonization occurring early in the period of secondary contact between *M. galloprovincialis* and *M. edulis* could have favoured the mixing of the different backgrounds over a large spatial scale. To illustrate this, we modified the initial conditions of the secondary contact model described above. The demes in the periphery of the range were fixed for one of the two backgrounds as previously; however each deme in the central portion of the chain of demes was initially assumed to be either fixed for *AB* or for

ab, randomly and independently drawn with equal probabilities. This situation simulates a random colonization of habitat patches by long-distance migrants (instead of a continuous, deterministic diffusion process) when the two entities came into contact. Polymorphism at the exogenous locus (adaptation to local variations in habitat) occurred within the two endogenous backgrounds. Figure 5 shows the result of one representative simulation that resulted in a two-scale mosaic hybrid zone. In the central portion of the chain of demes, the system formed multiple stable mosaic hybrid zones, which varied according to initial conditions, migration, selection, linkage disequilibria and habitat heterogeneities. In some places, linkage disequilibria between endogenous and exogenous loci broke down and local-adaptation genes segregate in a fixed endogenous background, while in other places linkage disequilibria were maintained and the coupling operated between endogenous loci and environmental variation (Fig. 5). We often obtained a two-scale mosaic structure with patches of parental endogenous backgrounds enclosed within the zone, in which habitat polymorphism segregates, and independent fine-scale mosaic hybrid zones, as in the simulation presented in Fig. 5 and in *Mytilus* hybrid zone in France. In the simulation chosen in Fig. 5, endogenous and exogenous

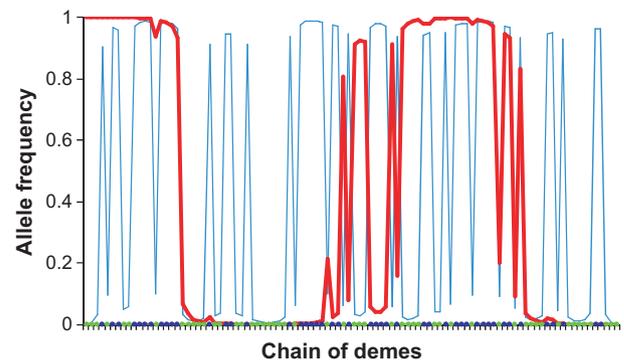


Fig. 5 Secondary contact between two intrinsically incompatible backgrounds with random colonization of habitat patches by long-distance migrants and polymorphism at the exogenous locus within each of the two endogenous backgrounds. Each deme in a central portion of 50 demes in a chain of 100 demes was randomly colonized by one alternative endogenous background. Each deme is assigned to habitat 1 (blue dots) or habitat 2 (green dots) at random. Selection is strong ($s = 0.9$, $t = 0.9$). Migration connects demes separated by up to six demes (see text). The system forms multiple stable mosaic hybrid zones. In some places, linkage disequilibria between endogenous and exogenous loci break down and local-adaptation genes segregate in a single endogenous background, while in other places linkage disequilibria are maintained and the coupling operates between endogenous loci and environmental variation.

loci phased in opposite directions in the central and peripheral zones. This was not the case in every simulation, but it can occur when habitat polymorphism was present in both endogenous backgrounds before the contact. To date, the GEA observed in the hybrid zone between *Mytilus edulis* and *M. galloprovincialis* has always been in the same direction (*M. edulis*-sheltered habitats/*M. galloprovincialis*-exposed habitats), but it does not mean reversed associations could not be found with further examination. Interestingly, we recently have sampled *M. galloprovincialis* mussels in a sheltered brackish habitat in the port of Cherbourg (Normandie, France), in a poorly known portion of the hybrid zone. More convincing examples of reversed GEAs will be given in the following section.

The coupling between an endogenous and an exogenous background therefore explains why GEAs are often observed at a local scale. GEAs are captured by neutral markers only in places where the coupling process can occur because habitats and gene pools are sufficiently intermingled to maintain linkage disequilibria between genetic incompatibilities, local-adaptation genes and neutral loci. The coupling model also predicts that if secondary contacts could occur several times, as in computer simulations, the phase of the association between the exogenous and the endogenous backgrounds could form in opposite directions (Fig. 4). Gould (1989) has argued that if we could replay the history of life it would certainly turn out to be different; in these fictive worlds, *Mytilus edulis* would perhaps seem to be adapted to exposed habitats and *M. galloprovincialis* to sheltered habitats, *Bombina variegata* would appear adapted to ponds and *B. bombina* to puddles and *Gryllus pennsylvanicus* would seem adapted to sand soils and *G. firmus* to loam soils. Unfortunately, secondary contacts often occur only once in nature.

Case studies of reversed or alternative GEAs

An attractive prediction of the coupling model is that the phase of the association between endogenous and exogenous backgrounds is not always constrained. If contacts between populations of the same species or subspecies occur in different locations, and if the initial conditions vary appropriately, then the same set of endogenous loci can couple with different sets of exogenous loci in the different locations. We have found some valuable examples of such varied GEAs at remote locations that might be explained by this process.

Mytilus edulis and *M. trossulus* in Europe and in North America. *M. edulis* and *M. galloprovincialis* have met only once in Europe, however *M. edulis* and *M. trossulus* – the third species of the *M. edulis* complex – have proba-

bly met on two independent occasions: in Europe and in North America (Riginos & Cunningham 2005). As mentioned above, in Europe *M. trossulus* is found in the Baltic Sea, in some fjords of Norway and bays of the Barents Sea. Consequently, *M. trossulus* has been inferred to be adapted to lower salinities (Johannesson *et al.* 1990; Riginos & Cunningham 2005). *M. trossulus* is also found in the western Atlantic where it forms another hybrid zone with *M. edulis*, which extends from Maine (Rawson *et al.* 2001) to Hudson Bay (Koehn *et al.* 1984). The spatial structure of the Western Atlantic hybrid zone between *M. edulis* and *M. trossulus* is a mosaic that resembles the European hybrid zone between *M. edulis* and *M. galloprovincialis*. It is probably a two-scale mosaic with enclosed patches of differentially introgressed parental backgrounds (Koehn *et al.* 1984) and several small-scale mosaic zones with environmental correlates (Bates & Innes 1995; Comesaña *et al.* 1999; Rawson *et al.* 2001; Toro *et al.* 2004). The spatial segregation strongly suggests that *M. edulis* is preferably found within bays and estuaries while *M. trossulus* is preferably found in oceanic coasts (Gartner-Kepkay *et al.* 1980; Gardner & Thompson 2001; Rawson *et al.* 2001). With respect to salinity gradients, the distribution in North America therefore appears to be the opposite of that observed in Europe. This led Riginos & Cunningham (2005) to suggest that local adaptation could have occurred after the secondary contacts. However marine mussels have always been confronted by salinity gradients, and so a more parsimonious explanation is that this reversed GEA provides an illustration of the prediction of the coupling model. The genetic integrity of the two species is mainly maintained by an efficient endogenous barrier. This has recently been verified through experimental demonstrations of gamete incompatibility (Rawson *et al.* 2003; Miranda *et al.* 2010) and hybrid inviability (Miranda *et al.* 2010). The coupling of the two endogenous backgrounds with habitat-adaptation genes could have occurred in opposite directions during two independent secondary contacts. It is also possible that opposite phasing could have occurred locally in some unexplored portions of these two large contact zones. Preliminary results from the Kola Bay in Russia indeed suggest such a complex relation with the environment, with inverted zonation at a small spatial scale (M.V. Katolikova, personal communication).

If our hypothesis proves to be true, it would mean that local adaptation has a negligible role in the maintenance of the genetic structure at neutral loci and of barriers to gene flow in *Mytilus* but is simply revealed within mosaic hybrid zones because in these zones habitat-adaptation genes are in linkage disequilibrium with endogenous backgrounds and neutral markers.

Semibalanus balanoides in Maine, the Gulf of St. Lawrence and Rhode Island. In the barnacle *S. balanoides*, allele frequencies at the mannose-6-phosphate isomerase (Mpi) locus are correlated with the degree of physiological stress experienced between high-tide ('hot') and low-tide ('cold') microhabitats (Schmidt & Rand 1999). In Maine (USA), genetic homogeneity was observed both among stages and habitats for two control loci (Gpi and mtDNA), while genetic differentiation appeared at the Mpi locus at the juvenile stage, reflecting habitat-specific differential mortalities (Schmidt & Rand 2001). This species therefore provides a beautiful example of local selection in a fine-grained environment. However, things might be more complex than selection on a single locus in a heterogeneous environment (Bierne *et al.* 2003a). Further north, in the Gulf of St. Lawrence, a genetic discontinuity is observed over a short distance at the position of the Miramichi River, not only at the Mpi locus but also at the Gpi locus (Holm & Bourget 1994; Véliz *et al.* 2004) and at two out of six microsatellites (Dufresne *et al.* 2002). The differential in Mpi allele frequency between the two sides of the Miramichi River is exactly the same as the difference between the two microhabitats in the Maine – the frequency of the F allele changes from ~50% (low-tide, south of Miramichi) to ~75% (high-tide, north of Miramichi). In such a large-scale spatial context it is difficult to understand why the allele frequency differential does not reach stronger values if selection directly affects allozyme loci. Furthermore, the ecological differences between each side of the Miramichi River are unclear. The north is slightly cooler (Drouin *et al.* 2002) while the frequency of the F allele, inferred to be adapted to higher temperature in the Maine, increase in the north, suggesting the relationship with temperature is possibly inverted. The Miramichi estuary, on the other hand, is likely to act as a natural barrier to dispersal (Drouin *et al.* 2002) capable to trap tension zones. Finally, Rand *et al.* (2002) replicated the study of GEA at the Mpi and Gpi loci further in the South in Narragansett Bay, Rhode Island. Not only did Rand *et al.* (2002) find a significant microhabitat zonation at the Gpi locus, but they also observed an opposite zonation pattern at the Mpi locus.

We propose that Gpi, Mpi and structured microsatellite loci might be simple markers of differentiated, partly incompatible, endogenous backgrounds. Interestingly, populations south to the Miramichi River have a similar genetic composition to European populations at the Mpi and Gpi loci (Holm & Bourget 1994), an observation which could have filled in a secondary contact scenario. Mpi and Gpi may possibly be directly involved in the barrier (Flight *et al.* 2010) but irrespectively of the environment (i.e. endogenous selection) and with a complex determinism, as differential fixation

is not observed either side of the Miramichi River. The cohesiveness of the two endogenous backgrounds as well as their associations with habitat-adaptation genes would vary between localities. Under this hypothesis, the reversed GEA observed between Maine and Rhode Island, or between Maine and the Miramichi River area is not surprising as it is one prediction of the coupling model.

Anopheles gambiae s.s. in Western and Central Africa. The malaria mosquito *A. gambiae sensu stricto*, which is the nominal species of the complex *A. gambiae sensu lato*, provides a convincing example of how different backgrounds, probably involved in different components of reproductive isolation (i.e. endogenous vs. exogenous), can form different associations at different places. In this case, both backgrounds have been characterized at the molecular level.

Chromosome surveys of inversion polymorphisms had led to the subdivision of *A. gambiae* s.s. into five chromosomal forms: Savanna, Mopti, Forest, Bamako and Bissau (Coluzzi *et al.* 1985; Touré *et al.* 1998). Cytogenetic studies indicated strong deviations from Hardy-Weinberg and linkage equilibrium and revealed association between inversion polymorphisms and environmental variation, such as the degree of aridity (Bryan *et al.* 1982; Coluzzi *et al.* 1985; Touré *et al.* 1998; Powell *et al.* 1999).

Extensive molecular analyses attempted to further distinguish the number of isolated or semi-isolated gene pools that exist in *A. gambiae* s.s., but demonstrated the existence of only two different entities, now referred to as molecular forms M and S (della Torre *et al.* 2001; Lehmann & Diabate 2008). The level of differentiation between the two forms is mainly restricted to small chromosome regions near the centromeres (Turner *et al.* 2005; White *et al.* 2010), including the X pericentric region to which the rDNA locus belongs; alleles at this locus are used to define the M and S forms (della Torre *et al.* 2005). Furthermore, the centromeric regions of all three chromosomes are in near-maximal linkage disequilibrium (White *et al.* 2010). Elsewhere in the genome, the level of genetic differentiation between the M and S forms is low though still significantly higher than between geographically distant populations of the same form (Wondji *et al.* 2002). It remains unclear whether the M/S barrier is semi-permeable – in this case chromosomal regions with high M/S differentiation correspond to the localization of genes involved in reproductive isolation – or whether the barrier is genome-wide – in this case differentiated regions correspond to regions of low recombination that differentiate faster owing to a stronger impact of hitchhiking and background selection (Noor & Bennett 2009; White *et al.* 2010).

However, the relationship between the molecular (M/S) and chromosomal forms is not simple. In West Africa (Mali and Burkina Faso), exact correspondence between molecular and chromosomal forms has been found: all Mopti individuals belong to the M molecular form while Savanna and Bamako chromosomal forms correspond to the S molecular form (this explains the origin of the M/S nomenclature). But this association breaks down in other areas of Africa (della Torre *et al.* 2001; Simard *et al.* 2009). In Cameroon (Central Africa), chromosomal arrangements assort independently of the M/S molecular forms (Simard *et al.* 2009). Each molecular form is polymorphic for chromosomal arrangements that approximately correspond to the Forest and Savanna chromosomal forms.

An abundant literature has emphasized the importance of chromosomal inversions in ecological adaptation (Powell *et al.* 1999). For instance, the frequency of the 2La arrangement correlates well with the degree of aridity in West African populations (Coluzzi 1992). Simard *et al.* (2009) recently studied GEA with both cytological and molecular markers and found that alternative homokaryotypes segregated in contrasting environments within both the M and S genetic backgrounds in Cameroon. Therefore inversion polymorphisms are probably directly involved in environmental adaptation. In contrast, the M and S molecular forms, the divergence of which has been studied in a chromosomally homosequential context (Turner *et al.* 2005), could reflect an endogenous barrier to gene flow, most probably pre-zygotic (White *et al.* 2010).

In this example the endogenous backgrounds (M/S) have become tightly coupled to an exogenous polymorphism (chromosomal arrangement) in one location, Western Africa, while exogenous loci have remained polymorphic within each endogenous background in another location, Central Africa. As a consequence a GEA involving variations in aridity is detected for the M/S polymorphisms in Western Africa while it is not detected, or with different environmental factors, in Central Africa (Simard *et al.* 2009). This suggests that local adaptation has a negligible effect on gene flow compared to the endogenous barrier. Interestingly, the chromosome 2 inversion that contributes in the adaptation of the M form to drier conditions than the S form in Western Africa is likely to have introgressed adaptively from *A. arabiensis*, another sister species of the complex (Besansky *et al.* 2003). Adaptive introgression of locally adapted genes through endogenous backgrounds is another prediction of the model.

This might be a case where the distinction between endogenous and exogenous selection is probably not clear-cut. For example, inversions might also have some slight intrinsic disadvantage as heterozygotes.

Nevertheless, discordance between karyotypic and molecular markers in *Anopheles gambiae* (as with the Australian Morabine grasshoppers of the genus *Vandiemenella*, Kawakami *et al.* 2009; Kearney & Hewitt 2009) illustrates how co-adapted alleles can couple at some locations, and scatter at others depending of the landscape, the genetic determinism and historical contingency.

Littorina saxatilis in Spain, Great Britain and Sweden. The intertidal snail *L. saxatilis* has become a popular model in the study of parallel ecological speciation (Rolán-Alvarez *et al.* 2004; Sadedin *et al.* 2009; Johannesson *et al.* 2010). *L. saxatilis* is direct-developing, lacking a dispersive larval stage, which could facilitate local adaptation. It inhabits the intertidal rocky shores throughout Europe. Two morphologically different ecotypes are usually found on the shore: a small, thin-shelled, morph with a wide aperture and a larger, thicker-shelled morph with a narrower aperture (Butlin *et al.* 2008). Pairs of divergent morphs have been well studied in three different areas: Galicia (northwest Spain), Yorkshire (northeast England), and around the Tjärnö marine biological laboratory in southwest Sweden. Within each site one morph occupies the high shore and the other morph the lower shore. The two morphs are separated by very narrow transition zones, sometimes only 2 m wide (Grahame *et al.* 2006). The two morphs are isolated by multiple mechanisms of reproductive isolation: local adaptation (Janson & Sundberg 1983; Rolán-Alvarez *et al.* 1997) and habitat preference (Cruz *et al.* 2004b) as well as assortative mating (Johannesson *et al.* 1995b; Hull 1998; Pickles & Grahame 1999; Cruz *et al.* 2004a; Hollander *et al.* 2005) and hybrid fitness depression (Hull *et al.* 1996; Cruz & Garcia 2001; Rolán-Alvarez 2007). The level of genetic differentiation between the two morphs is globally low but usually higher than between populations of the same morph separated by a similar distance (Grahame *et al.* 2006; Panova *et al.* 2006). However, genome scans have revealed that around 5% of the AFLP markers studied are high-Fst outliers (Wilding *et al.* 2001; Galindo *et al.* 2009). There are nonetheless very few convincing arguments to support the claim that these outlier loci are more affected by disruptive local adaptation than by hybrid fitness depression or assortative mating genes.

An interesting result is the opposite vertical zonation observed in Spain and England: in Spain, the large and thick morph occupies the high shore and the small and thin morph occupies the lower shore, while it is the reverse in England (Butlin *et al.* 2008). One explanation is that the environment to which morphs are differentially adapted also shows an opposite vertical relationship (Butlin *et al.* 2008). A thick shell with a narrow aperture is thought to improve resistance to crab predation, while a

small size and a large aperture, and thus a larger foot, would be an adaptation to withstand wave action. However, the two morphs have also been assumed to respond to other environmental gradients, such as emersion times (desiccation), salinity or temperature (Johannesson *et al.* 1993; Rolán-Alvarez *et al.* 1997). In the English shore as in many other shores, the upper part is exposed to stronger wave action and crabs are more abundant in the mid and low shore. In the Spanish shore, the risk of crab predation would possibly be higher in the upper shore and wave action stronger in the lower shore (Butlin *et al.* 2008), while the possible reversal of other intertidal gradients is unclear. The coupling model shows that one does not necessarily need to find a reverse relationship between the vertical zonation and the environment to explain inverted GEA. It is also possible that exogenous and endogenous barriers have phased differently in the two areas.

We here leave aside the question of whether the various barriers evolved independently at different places (parallel speciation, Johannesson *et al.* 2010) or whether they have a common origin and subsequently self-organized according to the history of population displacements and to the landscape while leaving the neutral diversity freed from equilibrating with geographical distance (Grahame *et al.* 2006). This issue may be tackled in the near future through detailed studies of sequence variation at outlier loci found in genome scans. If the divergence between alleles at these loci turns out to be very ancient, as suggested by preliminary results (Wood *et al.* 2008) or in another case study of parallel evolution in sticklebacks (Colosimo *et al.* 2005), the hypothesis of a common origin will become plausible, although alternative explanations implying evolution from standing genetic variation exist (Schluter & Conte 2009; Johannesson *et al.* 2010).

Ostrinia nubilalis in France, United States and Japan. *O. nubilalis* (corn borer) is a phytophagous insect that exhibits interesting patterns of association between genetic and environmental differentiation. Two pheromonal races, named E and Z according to the composition of the volatile compounds emitted during the mating process, have been described as reproductively isolated in sympatry (Malausa *et al.* 2005). In France, the E and Z races appear specialized to two different plant hosts: the E race is found on mugwort or hop, while the Z race is found on maize (Pelozuelo *et al.* 2004). In North America, where the European corn borer has been introduced at the beginning of the 20th century, the two pheromonal races coexist on a single host, maize. Furthermore, a voltinism polymorphism has evolved. *O. nubilalis* moths are either univoltine (one reproduction cycle per year, hereafter named

U strain) or bivoltine (two or more reproduction cycles per year, hereafter named B strain). This character can be considered either as an endogenous isolation system (moths are not emerging at the same time) or as an adaptation to a temperature gradient (moths living in the north of their distribution in the US are generally univoltine, while those living in the south tend to be bivoltine). Interestingly, in the US, an association can be found between the E/Z and U/B polymorphism. Moths can be BE, BZ or UZ, but UE phenotypes have never been observed (Dopman *et al.* 2010). This situation illustrates how the E/Z polymorphism, responsible for an endogenous barrier that was pre-existing in European populations, has coupled with an exogenous barrier associated with host choice in France, while it seems to have entered into a coupling process with an alternative environmental factor (temperature variation) in the US. Interestingly, in Japan the E/Z polymorphism exists in two different *Ostrinia* species, not only *O. nubilalis* but also *O. scapularis* (Huang *et al.* 2002). However, in Japan, *O. nubilalis* exclusively lives on maize while *O. scapularis* lives on hop and other dicotyledonous plants. This observation led Frolov *et al.* (2007) to revise the systematics of *Ostrinia* moths and to propose that moths living on maize should be considered as belonging to *O. nubilalis*, while those living on dicots should be named *O. scapularis*, including the E races initially named *O. nubilalis* in France. This illustrated that the interaction of multiple genetic barriers (pheromones, local adaptation to host plants, voltinism) can be so complex as to create a taxonomical conundrum even in well-studied taxa. Whatever the species names, the three geographic locations (France, US and Japan) clearly display alternative associations between an endogenous barrier (mate preference) and exogenous selection (adaptation to host plants or temperature). A broader analysis of European populations on a larger spatial scale should reveal whether the coupling between the E/Z backgrounds and the mugwort/maize environmental variation is only local (to France) or is more widespread.

Discussion

An alternative interpretation of GEAs

The approach of scanning genomes for loci with anomalously high levels of differentiation has become a standard of population genetics (Luikart *et al.* 2003). Sophisticated tests are continuously developed to identify so-called Fst outlier loci (Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009; Bazin *et al.* 2010). However, the exact form of the selection responsible for extreme differentiation is hardly ever

addressed and loci of greatly increased differentiation are simply assumed to be under divergent selection ('local adaptation'). This interpretation is sometimes given even in the absence of an observed relationship with an ecological variable. GEA is nonetheless often observed at outlier loci, which is then taken as additional support for the action of ecological selection either on the locus itself or on a linked locus. We argue here that GEA provides evidence that local adaptation exists somewhere in the genome but provides little indication that the identified loci (or small chromosomal regions surrounding them) directly respond to that selection. This is particularly true when outlier loci are found to represent a substantial portion of the panel of loci screened.

Decades of research on hybrid zones should have taught us that efficient genetic barriers to gene flow are very often endogenous – i.e. tension zones (Barton & Hewitt 1985). The existence of an endogenous genetic barrier is strongly suggested when many loci exhibit a concordant genetic structure. However, when GEA is observed it is very tempting to attribute the structure to environmental selection, one Holy Grail of evolutionary genetics. We believe genetic structure of this kind might often be due to many endogenous loci trapped at an environmental boundary by a smaller number of exogenous loci. Not only is this possibility theoretically expected but we have found examples that can plausibly be explained by this process. In the mosquito *A. gambiae s.s.*, the strength of the barrier to gene flow between the M and S forms is as strong in Cameroon, where endogenous backgrounds assort independently of the exogenous backgrounds, as in Mali where the endogenous and exogenous backgrounds have become coupled together (Wondji *et al.* 2002; White *et al.* 2010). In *Mytilus* mussels, habitat adaptation polymorphisms are likely to segregate within monospecific patches of populations but have remained invisible to molecular studies to date. Their existence is attested by local GEAs observed in fine-grained hybrid zones, in which coupling occurred with the endogenous barrier, and by the opposite relationship to environmental heterogeneity in two replicated secondary contact zones in Europe and North America.

When the concomitant action of endogenous and exogenous factors is recognized as contributing to reproductive isolation, either a secondary contact is suspected and one refers to the hybrid zone framework, or the secondary contact scenario is thought unlikely (albeit difficult to rule out definitively) and one refers to ecological speciation. In the latter case, exogenous selection is seen as the driving force that led to the evolution of reproductive isolation, the first spark to initiation of the subsequent accumulation of barrier genes (Nosil *et al.* 2009; Schluter & Conte 2009; Via 2009), possibly in the chromosomal

neighbourhood of locally adapted genes (Smadja *et al.* 2008; Via & West 2008). This review is not the place for a discussion of the importance of ecology-driven divergent selection in speciation – a complex issue that despite abundant attention, and much speculation, is not yet settled. There is nonetheless a point that is relevant here: the ecological speciation hypothesis faces a 'chicken-egg dilemma', but typically assumes that the exogenous barrier has arisen first and then drives the evolution of the endogenous barrier. While this scenario is attractive, the coupling model presented here provides an alternative possibility. New environmental adaptation can trap a pre-existing endogenous barrier from a distant location (Fig. 2), and most species are likely to have endogenous barriers somewhere in their distribution range, either because of a history of vicariance (Hewitt 2000) or because genetic incompatibilities accumulate in parapatry (Kondrashov 2003; Navarro & Barton 2003; Gavrilets 2004). New adaptations are not required to drive the evolution of a new barrier; they can couple with a pre-existing one. One possible example of this process is found in the corn borer, *Ostrinia nubilalis*, which colonized maize after its introduction into Europe, around 500 years ago. Adaptation to maize has probably resulted in the capture of a pre-existing endogenous barrier between *O. nubilalis* and *O. scapularis* rather than the evolution of a new one through the process of ecological speciation. Similarly, the complexity and the genomic extent of the barrier to gene flow between the hawthorn and apple host races of the maggot fly *Rhagoletis pomonella* (Michel *et al.* 2010) does not suggest an origin as recent as 150 years – the estimated age of the host shift to apple. When exogenous and endogenous barriers arise together, it is not necessarily meaningful to ask which of the two drove the evolution of the other. In a groundbreaking mathematical analysis, Barton & de Cara (2009) recently revised our view of the reinforcement process: strong isolation can evolve through the coupling of any kind of incompatibility, whether pre- or post-zygotic. The coupling occurs when selection for increased variance in heterozygosity drives the increase of linkage disequilibrium, which couples different components of reproductive isolation. That is why Barton & de Cara (2009) argued that the action of selection on linkage disequilibria is adaptive. Although these authors considered intrinsic incompatibilities only, disruptive exogenous selection often favours maximal variance and exogenous loci can also enter the coupling process, as shown in our model. By presenting speciation as a coupling process, Barton and de Cara (also see Udovic 1980; Kirkpatrick & Ravigné 2001) reconciled traditionally opposed views of speciation: the evolution of reproductive evolution is seen as an accumulation of incompatibilities of any kind, pre- or post-zygotic, endogenous or exogenous, regardless of their nature or

order of appearance. Acknowledging that speciation is a gradual multifactorial process renders less important the questions of which type of factor acts first or most strongly, questions to which a general answer may not exist.

Why are endogenous barriers underappreciated to explain GEAs?

It is difficult to find evidence for endogenous selection. Our own experience in the study of hybrid fitness led us to realize how difficult it can sometimes be to demonstrate the existence and causes of hybrid fitness depression. Two decades of research on the hybrid zone between the mussel species *M. edulis* and *M. galloprovincialis* failed to demonstrate lower fitness of hybrids (Wilhelm & Hilbish 1998). On the other hand, associations between genetic structure and environmental factors (salinity, wave exposure), led researchers to emphasize habitat specialization as the principal mechanism of reproductive isolation between the two species (Gardner 1994). However, the pattern of genetic structure within hybrid populations suggested exogenous selection could not be acting alone and that endogenous isolation mechanisms must also have existed (Bierne *et al.* 2002b). To uncover hybrid fitness depression, we had to perform controlled crosses in the lab, and to go through two generations of hybridization (Bierne *et al.* 2006). Unfit hybrid genotypes, those key genotypes that prevent introgression (Barton 2001), are expected to be rare in nature. In order to estimate their fitness accurately, they must be produced in large numbers through experimental crossing. In addition, hybrid dysgenesis often appears only in the F₂s (after one generation of recombination) while heterosis is often observed in F₁ progeny (Dobzhansky 1952; Alibert *et al.* 1997; Edmands 1999; Bierne *et al.* 2002a). We believe the *Mytilus* case is not isolated because controlled crosses are not always feasible and the F₂ generation rarely investigated. Furthermore, hybrid fitness is often investigated in nature through the study of associations between genotypes at marker loci and phenotypes. However, 'hybrid individuals' sampled in natural populations and used to estimate hybrid fitness are often complex genetic mosaics in which neutral loci have a loose linkage disequilibrium with barrier loci. Although the study of genotype/phenotype associations can often be powerful at inferring phenotypic differences between parental backgrounds in hybrid zones, its efficiency in accurately inferring the fitness of 'hybrids', an ambiguously defined genotypic category, is less clear (Boecklen & Howard 1997).

Identifying genetic incompatibilities is even more difficult. For instance, the identification of barrier genes

is a recent development in the study of speciation that awaited the development of molecular techniques in model species (Noor & Feder 2006; Wolf *et al.* 2010). The genetics of endogenous post-zygotic isolation has long been mostly investigated in *Drosophila*, focussing on X-linked genes with large effects on fitness and between well-delimited species (Orr *et al.* 2004). Decades of research pioneered by Dobzhansky provided overwhelming evidence that hybrid fitness depression results from the accumulation of DM incompatibilities (Coyne & Orr 2004). However, despite valuable recent works in other species (Presgraves 2010) the study of more loosely isolated backgrounds and of autosomal genes with moderate effects on fitness remains limited (Rieseberg & Buerkle 2002; Bierne *et al.* 2006; Edmands *et al.* 2009; Wolf *et al.* 2010). In *Mytilus* mussels, we were lucky enough that the number of DM incompatibilities was sufficiently high for our markers to easily map them, but it proved more difficult to infer the genetic determination of hybrid dysgenesis in *Tigriopus* copepods (Edmands *et al.* 2009). For weaker barriers involving fewer genetic incompatibilities, the effort needed to map interacting incompatibilities is much greater.

It is usually thought that there exists an optimal genetic distance above which the beneficial effect of hybridization (i.e. heterosis) are overwhelmed by its negative effects (i.e. outbreeding depression, Waser 1993). Surveys that have tried to identify this optimum have often found the distance to be very small (Escobar *et al.* 2008). At larger scales, the correlation between parental divergence and post-zygotic isolation is positive (Edmands 1999, 2002), although the scaling remains somewhat blurred because the metrics used are either spatial, ranging from metres to thousands of kilometres, or temporal, ranging from thousands to millions of years (Edmands 2002). In any case, hybrid fitness depression is often observed at a spatial scale for which neutral markers do not exhibit a strong genetic differentiation (Escobar *et al.* 2008).

In the era of genomics, research on the genetic basis of reproductive isolation is likely to reveal the importance of DM incompatibilities even between populations within species, provided it is looked for and inherent analytical difficulties are solved. Recent studies are progressing in this direction. Studying segregation distortion in an F₂ cross between two divergent populations of *Mimulus guttatus*, Hall & Willis (2005) found that half of all markers significantly departed from Mendelian expectations, allowing them to map 12 genetic incompatibilities. A similar level of departure was detected in an interspecific map between *M. guttatus* and *M. nasutus*. Similarly, McDaniel *et al.* (2007) found a high rate of segregation distortion in an interpopulation cross of the

moss *Ceratodon purpureus*, with evidence that this arises from epistatic interactions. Montooth *et al.* (2010) recently studied mitochondrial-nuclear incompatibilities in flies by combining mitochondria with either interspecific nuclear backgrounds or with a nuclear background of a different population of the same species. Significant epistasis for male fitness was observed in an intraspecific cross, and surprisingly this was the strongest effect, stronger than in an interspecific combination. Lachance & True (2010) studied epistatic fitness interactions between the X chromosome and autosomal genetic backgrounds derived from different geographic locations of *Drosophila melanogaster* and found considerable amounts of recessive incompatibilities. Recently, genetic incompatibilities have been identified between *Arabidopsis thaliana* accessions (Bikard *et al.* 2009; Alcazar *et al.* 2010). All of these studies suggest that DM incompatibilities are widespread, even within a species, although their characterization is difficult and time-consuming. There is thus no reason to dismiss the possibility of such incompatibilities in the interpretation of Fst scans. Interestingly, the segregation of genetic incompatibilities within *Drosophila* and *Arabidopsis* have been observed at spatio-temporal scales which are much smaller than usually anticipated (Kolaczowski *et al.* 2010; Turner *et al.* 2010).

It is not appreciated that trapping of tension zones can occur. It is well recognized that tension zones are expected to become trapped by natural barriers to dispersal (Barton 1979a; Barton & Hewitt 1985; Hewitt 1988). Despite this prediction, natural barriers are still often interpreted as explaining not only the position of genetic breaks, but also the genetic breaks themselves. Even worse, when a natural barrier coincides with an environmental boundary, the genetic structure is often said to result from local adaptation. Although well established, Barton's trapping hypothesis does seem to us to be insufficiently appreciated outside of the hybrid zone literature. The expectation that tension zones can coincide with environmental boundaries, even without a barrier to dispersal, remains implicit even in much of the hybrid zone literature so it is not surprising that it is not appreciated elsewhere. However, it has long been known that tension zones are expected to stabilize where parental fitnesses become equal and in such a way as to minimize their length (Barton & Hewitt 1985), and it should be an intuitive outcome of speciation and cline theories that endogenous and exogenous backgrounds can couple together. It is therefore timely to emphasize that tension zones can be trapped by exogenous clines at ecotones, and more generally to encourage molecular ecologists to disentangle the question of the cause of the position of a genetic break/cline

and the question of its nature. An interesting issue is how natural barriers compete with environmental variation in attracting tension zones. Tension zones can be trapped by minor local barriers and will not easily get to an environmental transition. It is however plausible that a continual process of extinction and recolonization could tend to bring exogenous and endogenous clines together, and that once this happens, the strong net effect of environment-related selection may push hybrid zones to the ecotone more effectively. More generally, random variations in population density can help to regularly relocate endogenous clines until overlapping with an exogenous cline. Further modelling is required to investigate this issue.

Ideas from the hybrid zone literature have not been considered in the Fst outlier literature. The hybrid zone literature is consistently ignored in the literature on local adaptation at the molecular level and in Fst scan surveys. The reason is probably a mistaken belief that the cases are quite different. We believe they are in fact the same and that the population genetics of local adaptation should consider arguments from the hybrid zone literature more seriously.

In the hybrid zone literature of the 1980s, while genetic markers were mostly allozymes and the number of polymorphic loci available could not exceed a few dozen, it was common practice to identify a handful of loci that best discriminated partially isolated forms. After preliminary analyses of genetic differentiation, panels of so-called 'diagnostic' or 'semi-diagnostic' markers were consistently used in further field studies and lab experiments ignoring less informative loci. For example, this was true of *Mytilus* hybrid zones that were consistently analysed with four or five allozymes chosen from a possible 24 (Bierne *et al.* 2003d) and regrettably are now often analysed with a single DNA marker only. This was also true for other famous hybrid zones such as the *Bombina* toads hybrid zone studied with five allozymes from a possible 29 (Szymura 1993), the *Gryllus* crickets hybrid zone studied with three allozymes from a possible 23 (Harrison 1979), or the *Mus* mice hybrid zone studied with 10 allozymes from a possible 36 (Bonhomme *et al.* 1984; Raufaste *et al.* 2005). Although these markers were recognized as being affected by reproductive isolation, potentially because they were directly involved, they were used as markers of the genetic backgrounds responsible for the 'superphenotype' that combined the effect of every reproductive isolation factor. The initial scan of differentiation, although modest, resembles the Fst scan strategy that is increasingly used nowadays (Luikart *et al.* 2003; Beaumont 2005; Storz 2005; Nosil *et al.* 2009). Strangely enough, the recent genome scan literature

contains a far narrower range of hypotheses than the older hybrid zone literature for interpreting high F_{st} outliers, almost always invoking local selection. Such an interpretation avoids the questions of whether selection is acting now or acted in the past, whether selection affects, or has affected, the focal loci directly or indirectly, whether the association with a trait or an environment is direct or indirect, whether it is selection or other factors that similarly modifies genotypic frequencies (e.g. pre-zygotic isolation mechanisms, habitat choice) and finally consistently ignores the possible existence of hybrid fitness depression. One might argue that we are comparing different time scales and differentiation levels. However, F_{st} scans have often been conducted between recognized isolated forms that would fit well in the hybrid zone framework (Wilding *et al.* 2001; Murray & Hare 2006; Via & West 2008; Nielsen *et al.* 2009; Michel *et al.* 2010) and conversely, some hybrid zones display very low levels of genetic differentiation at most loci (Halliday *et al.* 1983; Nielsen *et al.* 2003). Barton & Hewitt (1985) estimated that on average, 14% of loci showed clear genetic differences in the 34 hybrid zones with molecular data they reviewed. This is not very different from the proportion of F_{st} outlier loci usually detected in modern genome scans. Finally, few species are likely to have remained unperturbed by a history of population displacement, fragmentation and secondary contact (Avice 2000; Hewitt 2000; Lewontin 2002), providing ample opportunities for the evolution of partially isolated genetic backgrounds. Genetic barriers and hybrid zones are expected to be ubiquitous. The question is less their existence than their intensities, and the number of loci needed to reveal partially isolated backgrounds of various natures. The detailed study of the spatial population genetics of outlier loci identified from genome scans is likely to reveal that multiple outliers often display concordant geographic structure, as is observed in the first few examples (Coop *et al.* 2009; Bradbury *et al.* 2010; Hohenlohe *et al.* 2010). Geographic coincidence calls for nontrivial explanations – it requires either that all the relevant environmental variables are concentrated at the same location or more likely that incompatible endogenous backgrounds largely contribute to the genetic structure observed in natural populations (Barton & Hewitt 1985).

Conclusion

The early study of hybrid zones was arguably characterized by excessive optimism about our ability to infer evolutionary process from genetic patterns. A few decades of research later these hopes have been tempered, and it is widely accepted that long term experiments in

the lab and field must accompany inferences from genetic data (Moore & Price 1993; Harrison 1998; Ross & Harrison 2002). However, the same excessive optimism now seems to underlie the genome scan approach, which mostly focuses on a single concept – local adaptation – to explain unusual differentiation, and seems to ignore the complexities of (i) population history, (ii) the genetic architecture of barriers to gene flow, (iii) the interaction between selected loci of different types, and (iv) the indirect path through which selection affects neutral variation. Our understanding of the genetics of local adaptation must be informed by arguments from the hybrid zone literature because the two biological situations differ in respect of the intensity of the barrier to gene flow but not necessarily in respect of its nature, which is often multifactorial, and both exogenous and endogenous. Barriers of both types have a tendency to become coupled, such that tension zones can come to coincide with habitat boundaries, while not themselves involved in local adaptation.

Acknowledgements

The first author thanks Louis Bernatchez and members of his lab for a stimulating week in Québec which initiated this review. The authors thank Matthieu Faure, Pierre-Alexandre Gagnaire and Sergine Ponsard for insightful discussion as well as David Rand 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 2011-033 of Institut des Sciences de l'Évolution de Montpellier.

References

- Alcazar R, Garcia AV, Kronholm I *et al.* (2010) Natural variation at Strubbelig Receptor Kinase 3 drives immune-triggered incompatibilities between *Arabidopsis thaliana* accessions. *Nature Genetics*, **42**, 1135–1139.
- Alibert P, Fel-Clair F, Manolakou K, Britton-Davidian J, Auffray J-C (1997) Developmental stability, fitness, and trait size in laboratory hybrids between European subspecies of the house mouse. *Evolution*, **51**, 1284–1295.
- Arnold ML (1997) *Natural Hybridization and Evolution*. Oxford University Press, New York.
- Avice JC (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, **63**, 62–76.
- Avice JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.
- Bahri-Sfar L, Lemaire C, Ben Hassine OK, Bonhomme F (2000) Fragmentation of sea bass populations in the western and eastern Mediterranean as revealed by microsatellite polymorphism. *Proceedings – Royal Society of London. Biological Sciences*, **267**, 929–935.

- Barton NH (1979a) The dynamics of hybrid zone. *Heredity*, **43**, 341–359.
- Barton NH (1979b) Gene flow past a cline. *Heredity*, **43**, 333–339.
- Barton NH (1980) The hybrid sink effect. *Heredity*, **44**, 277–278.
- Barton NH (1983) Multilocus clines. *Evolution*, **37**, 454–471.
- Barton NH (1986) The effects of linkage and density-dependant regulation on gene flow. *Heredity*, **57**, 415–426.
- Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**, 1553–1562.
- Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.
- Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, **56**, 357–376.
- 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.
- Bates JA, Innes DJ (1995) Genetic variation among populations of *Mytilus ssp* in eastern Newfoundland. *Marine Biology*, **124**, 417–424.
- Bazin E, Dawson KJ, Beaumont MA (2010) Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, **185**, 587–602.
- Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, **20**, 435–440.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Bengtsson BO (1985) The flow of genes through a genetic barrier. In: *Evolution: Essays in Honor of John Maynard Smith* (eds Greenwood JJ, Harvey PH, Slatkin M), pp. 31–42. Cambridge University Press, Cambridge.
- Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1783–1800.
- Berry AJ, Ajioka JW, Kreitman M (1991) Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. *Genetics*, **129**, 1111–1117.
- Bert TM, Arnold WS (1995) An empirical test of predictions of two competing models for the maintenance and fate of hybrid zones: both models are supported in a hard-clam hybrid zone. *Evolution*, **49**, 276–289.
- Besansky NJ, Krzywinski J, Lehmann T *et al.* (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences, USA*, **100**, 10818–10823.
- Bidau CJ (1990) The complex Robertsonian system of *Dichroplus pratensis* (Melanoplinae, Acrididae) II. Effects of the fusion polymorphisms on chiasma frequency and distribution. *Heredity*, **64**, 145–159.
- Bierne N (2001) *Sélection et dispersion larvaire dans la zone d'hybridation des moules côtières Mytilus edulis et M. galloprovincialis*. Université de Montpellier II, Montpellier.
- Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, **64**, 3254–3272.
- Bierne N, David P, Boudry P, Bonhomme F (2002a) Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M-galloprovincialis*. *Evolution*, **56**, 292–298.
- Bierne N, David P, Langlade A, Bonhomme F (2002b) Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? *Marine Ecology-Progress Series*, **245**, 157–170.
- Bierne N, Lenormand T, Bonhomme F, David P (2002c) Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow? *Genetical Research*, **80**, 197–204.
- Bierne N, Bonhomme F, David P (2003a) Genetics at larval stage in marine bivalves. In: *Recent Advances in Marine Biotechnology 10* (eds Fingerma M, Nagabhushanam R), pp. 239–262. Science Publishers, Inc., Enfield, NH, USA/Plymouth, UK.
- Bierne N, Bonhomme F, David P (2003b) Habitat preference and the marine-speciation paradox. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1399–1406.
- Bierne N, Borsa P, Daguin C *et al.* (2003c) Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M-galloprovincialis*. *Molecular Ecology*, **12**, 447–461.
- Bierne N, Daguin C, Bonhomme F, David P, Borsa P (2003d) Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology*, **12**, 2505–2510.
- Bierne N, Bonhomme F, Boudry P, Szulkin M, David P (2006) Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1253–1260.
- Bikard D, Patel D, Le Métte C *et al.* (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science*, **323**, 623–626.
- Billard E, Serrão E, Pearson G, Destombe C, Valero M (2010) *Fucus vesiculosus* and *spiralis* species complex: a nested model of local adaptation at the shore level. *Marine Ecology Progress Series*, **405**, 163–174.
- Boecklen WJ, Howard DJ (1997) Genetic analysis of hybrid zones: numbers of markers and power of resolution. *Ecology*, **78**, 2611–2616.
- Bonhomme F, Catalan J, Britton-Davidian J *et al.* (1984) Biochemical diversity and evolution in the genus *Mus*. *Biochemical Genetics*, **22**, 275–303.
- Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.
- Bradbury IR, Hubert S, Higgins B *et al.* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3725–3734.
- Britton-Davidian J, Fel-Clair F, Lopez J, Alibert P, Boursot P (2005) Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. *Biological Journal of the Linnean Society*, **84**, 379–393.
- Bryan JH, Di Deco MA, Petrarca V, Coluzzi M (1982) Inversion polymorphism and incipient speciation in *Anopheles gambiae s. str.* in The Gambia, West Africa. *Genetica*, **59**, 167–176.
- Bufalova EN, Strelkov PP, Katolikova MV, Sukhotin AA, Kozin MB (2005) *Mytilus* of the Tuva Bay (Kola Bay, Barents Sea). *Vestnik Université Saint-Petersbourg*, **3**, 99–105.
- Bulnheim HP, Scholl A (1981) Genetic variation between geographic populations of the amphipods *Gammarus zaddachi* and *G. salinus*. *Marine Biology*, **64**, 105–115.

- Burton RS (1998) Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution*, **52**, 734–745.
- Butlin RK (1989) Reinforcement of premating isolation. In: *Speciation and Its Consequences* (eds Otte D, Endler JA), pp. 158–179. Sinauer, Sunderland, MA.
- Butlin RK, Galindo J, Grahame J (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B*, **363**, 2997–3007.
- Calcagno V, Thomas Y, Bourguet D (2001) Sympatric host races of the European corn borer: adaptation to host plants and hybrid performance. *Journal of Evolutionary Biology*, **20**, 1720–1729.
- Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.
- 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. *Genetic Research*, **70**, 155–174.
- Chou J-Y, Hung Y-S, Lin K-H, Lee H-Y, Leu J-Y (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biology*, **8**, e1000432.
- Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, **307**, 1928–1933.
- Coluzzi M (1992) Malaria vector analysis and control. *Parasitology Today*, **8**, 113–118.
- Coluzzi M, Petrarca V, Di Deco MA (1985) Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bollettino di Zoologia*, **52**, 45–63.
- Comesaña AS, Toro JE, Innes DJ, Thompson RJ (1999) A molecular approach to the ecology of a mussel (*Mytilus edulis*-*M. trossulus*) hybrid zone on the east coast of Newfoundland, Canada. *Marine Biology*, **133**, 213–221.
- Coop G, Pickrell JK, Novembre J *et al.* (2009) The role of geography in human adaptation. *PLoS Genetics*, **5**, e1000500.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, MA.
- Cruz R, Garcia C (2001) Disruptive selection on female reproductive characters in a hybrid zone of *Littorina saxatilis*. *Evolutionary Ecology*, **15**, 167–182.
- Cruz R, Carballo M, Conde-Padin P, Rolan-Alvarez E (2004a) Testing alternative models for sexual isolation in natural populations of *Littorina saxatilis*: indirect support for by-product ecological speciation? *Journal of Evolutionary Biology*, **17**, 288–293.
- Cruz R, Vilas C, Mosquera J, Garcia C (2004b) Relative contribution of dispersal and natural selection to the maintenance of a hybrid zone in *Littorina*. *Evolution*, **58**, 2734–2746.
- Cunningham CW, Collins TM (1994) Developing model systems for molecular biogeography: Vicariance and interchange in marine invertebrates. In: *Molecular Ecology and Evolution: Approaches and Applications* (eds Schierwater B, Streit B, Wagner GP, DeSalle R), pp. 405–433. Birkhäuser verlag, Basel.
- Daguin C (2000) *Phylogéographie des moules du complexe d'espèces Mytilus edulis Thèse de doctorat*. Université de Montpellier II, Montpellier.
- della Torre A, Fanello C, Akogbeto M *et al.* (2001) Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Molecular Biology*, **10**, 9–18.
- della Torre A, Tu Z, Petrarca V (2005) On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochemistry and Molecular Biology*, **35**, 755–769.
- Diabaté A, Dabire RK, Kim EH *et al.* (2005) Larval development of the molecular forms of *Anopheles gambiae* (Diptera: Culicidae) in different habitats: a transplantation experiment. *Journal of Medical Entomology*, **42**, 548–553.
- Dobzhansky T (1952) *Nature and Origin of Heterosis*. Iowa State College Press, New York.
- Dopman EB, Robbins PS, Seaman A (2010) Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution*, **64**, 881–902.
- Drouin C-A, Bourget E, Tremblay R (2002) Larval transport processes of barnacle larvae in the vicinity of the interface between two genetically different populations of *Semibalanus balanoides*. *Marine Ecology Progress Series*, **229**, 165–172.
- Dufresne F, Bourget E, Bernatchez L (2002) Differential patterns of spatial divergence in microsatellite and allozyme alleles: further evidence for locus-specific selection in the acorn barnacle, *Semibalanus balanoides*? *Molecular Ecology*, **11**, 113–123.
- Edmands S (1999) Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution*, **53**, 1757–1768.
- Edmands S (2002) Does parental divergence predict reproductive compatibility? *Trends in Ecology & Evolution*, **17**, 520–527.
- Edmands S, Northrup SL, Hwang AS (2009) Maladapted gene complexes within populations of the intertidal copepod *Tigriopus californicus*? *Evolution*, **63**, 2184–2192.
- Endler JA (1977) *Geographic Variation, Speciation and Clines*. Princeton University Press, Princeton, NJ.
- Escobar JS, Nicot A, David P (2008) The different sources of variation in inbreeding depression, heterosis and outbreeding depression in a metapopulation of *Physa acuta*. *Genetics*, **180**, 1593–1608.
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.
- Faure MF, David P, Bonhomme F, Bierne N (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evolutionary Biology*, **8**, 164.
- Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.
- Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics*, **78**, 737–756.
- Ferree PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biology*, **7**, e1000234.
- Flight PA, Schoepfer SD, Rand DM (2010) Physiological stress and the fitness effects of Mpi genotypes in the acorn barnacle *Semibalanus balanoides*. *Marine Ecology Progress Series*, **404**, 139–149.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

- Frolov AN, Bourguet D, Ponsard S (2007) Reconsidering the taxonomy of several *Ostrinia* species in the light of reproductive isolation: a tale for Ernst Mayr. *Biological Journal of the Linnean Society*, **91**, 49–72.
- Gaggiotti OE, Bekkevoold D, Jørgensen HBH *et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**, 2939–2951.
- Galindo J, Moran P, Rolan-Alvarez E (2009) Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology*, **18**, 919–930.
- 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.
- Gardner JPA (1994) The structure and dynamics of naturally occurring hybrid *Mytilus edulis* and *M. galloprovincialis* (Bivalvia, Mollusca) populations: review and interpretation. *Archiv für Hydrobiologie (supplement band)*, **99**, 37–71.
- Gardner JPA, Skibinski DOF (1990) Genotype-dependant fecundity and temporal variation of spawning in hybrid mussel populations. *Marine Biology*, **105**, 153–162.
- Gardner JPA, Thompson RJ (2001) The effects of coastal and estuarine conditions on the physiology and survivorship of the mussels *Mytilus edulis*, *M. trossulus* and their hybrids. *Journal of Experimental Marine Biology and Ecology*, **265**, 119–140.
- Gartner-Kepekay KE, Dickie LM, Freeman KR, Zouros E (1980) Genetic differences and environments of mussel populations in the maritime provinces. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 775–782.
- Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton and Oxford.
- Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences, USA*, **102**, 18040–18045.
- Gavrilets S, Li H, Vose MD (2000) Patterns of parapatric speciation. *Evolution*, **54**, 1126–1134.
- Gilg MR, Hilbish TJ (2003) Patterns of larval dispersal and their effect on the maintenance of a blue mussel hybrid zone in southwestern England. *Evolution*, **57**, 1061–1077.
- Gillespie JH (2000) Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics*, **155**, 909–919.
- Gould SJ (1989) *Wonderful Life: The Burgess Shale and the Nature of History*. Norton, New York.
- 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.
- Hall MC, Willis JH (2005) Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: implications for genomic divergence. *Genetics*, **170**, 375–386.
- Halliday RB, Barton NH, Hewitt GM (1983) Electrophoretic analysis of a chromosomal hybrid zone in the grasshopper *Podisma pedestris*. *Biological Journal of the Linnean Society*, **19**, 51–62.
- Harrison RG (1979) Speciation in north american field crickets: evidence from electrophoretic comparisons. *Evolution*, **33**, 1009–1023.
- Harrison RG (1983) Barriers to gene exchange between closely related cricket species. I. *Laboratory hybridization studies*. *Evolution*, **37**, 245–251.
- Harrison RG (1985) Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. *Evolution*, **39**, 244–259.
- Harrison RG (1986) Pattern and process in a narrow hybrid zone. *Heredity*, **56**, 337–349.
- Harrison RG (1993) Hybrids and hybrid zones: historical perspective. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 3–12. Oxford University Press, New York, Oxford.
- Harrison RG (1998) Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: *Endless Forms: Species and Speciation* (eds Howard DJ, Berlocher SH), pp. 19–31. Oxford University Press, NY and Oxford.
- Harrison RG, Rand DM (1989) Mosaic hybrid zone and the nature of species boundaries. In: *Speciation and its Consequences* (eds. Otte D, Endler JA), pp. 111–133. Sinauer Ass. Inc., Sunderland, MA.
- Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, **7**, 1–32.
- Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007a) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity*, **99**, 592–600.
- Hemmer-Hansen J, Nielsen EE, Gronkjaer P, Loeschcke V (2007b) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Molecular Ecology*, **16**, 3104–3118.
- Hewitt GM (1988) Hybrid zones, natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, **3**, 158–167.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195; discussion 195.
- Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: Geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.
- Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics*, **73**, 95–108.
- Hofman S, Spolsky C, Uzzell T *et al.* (2007) Phylogeography of the fire-bellied toads *Bombina*: independent Pleistocene histories inferred from mitochondrial genomes. *Molecular Ecology*, **16**, 2301–2316.
- Hohenlohe PLA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in three spine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

- Hollander J, Lindegarth M, Johannesson K (2005) Local adaptation but not geographical separation promotes assortative mating in a snail. *Animal Behaviour*, **70**, 1209–1219.
- Holm ER, Bourget E (1994) Selection and population genetic structure of the barnacle *Semibalanus balanoides* in the northwest Atlantic and gulf of St-Lawrence. *Marine Ecology Progress Series*, **113**, 247–256.
- Howard DJ, Waring GL (1991) Topographic diversity, zone width, and strength of reproductive isolation in a zone of overlap and hybridization. *Evolution*, **45**, 1120–1135.
- Huang Y, Takanashi T, Hoshizaki S, Tatsuki S, Ishikawa Y (2002) Female sex pheromone polymorphism in Adzuki Bean Borer, *Ostrinia scapulalis*, is similar to that in European Corn Borer, *O. nubilalis*. *Journal of Chemical Ecology*, **28**, 533–539.
- Hull SL (1998) Assortative mating between two morphs of *Littorina saxatilis* on a shore in Yorkshire. *Hydrobiologia*, **378**, 79–88.
- Hull SL, Grahame J, Mill PJ (1996) Morphological divergence and evidence for reproductive isolation in *Littorina saxatilis* (Olivi) in northeast England. *Journal of Molluscan Studies*, **62**, 89–99.
- Hunt WG, Selander RK (1973) Biochemical genetics of hybridisation in European house mice. *Heredity*, **31**, 11–33.
- Ingvarsson PK, Whitlock MC (2000) Heterosis increases the effective migration rate. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1321–1326.
- Janson K, Sundberg P (1983) Multivariate morphometric analysis of 2 varieties of *Littorina saxatilis* from the Swedish West-Coast. *Marine Biology*, **74**, 49–53.
- Johannesson K, Andre C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, **15**, 2013–2029.
- Johannesson K, Kautsky N, Tedengren M (1990) Genotypic and phenotypic differences between baltic and north sea populations of *M.edulis* evaluated through reciprocal transplantations. II Genetic variation. *Marine Ecology Progress Series*, **59**, 211–219.
- Johannesson K, Johannesson B, Rolan-Alvarez E (1993) Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail. *Evolution*, **47**, 1770–1787.
- Johannesson K, Johannesson B, Lundgren U (1995a) Strong natural selection causes microscale allozyme variation in a marine snail. *Proceedings of the National Academy of Sciences, USA*, **92**, 2602–2606.
- Johannesson K, Rolan-Alvarez E, Ekendahl A (1995b) Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution*, **49**, 1180–1190.
- Johannesson K, Panova M, Kempainen P *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.
- Kawakami T, Butlin RK, Adams M *et al.* (2009) Re-examination of a proposed case of stasipatric speciation: phylogeography of the Australian morabine grasshoppers (*Vandiemenella viatica* species group). *Molecular Ecology*, **18**, 3429–3442.
- Kearney M, Hewitt GM (2009) Stasipatric speciation: resurrecting a system to bury a hypothesis? *Molecular Ecology*, **18**, 3331–3333.
- Kirkpatrick M, Ravigné V (2001) Speciation by natural and sexual selection. *American Naturalist*, **159**, S22–S35.
- Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular and Biology Evolution*, **23**, 482–490.
- Koehn RK, Hilbish TJ (1987) The adaptive importance of genetic variation. *American Scientist*, **75**, 134–141.
- Koehn RK, Hall JG, Innes DJ, Zera AJ (1984) Genetic differentiation of *Mytilus edulis* in eastern North America. *Marine Biology*, **79**, 117–126.
- Kolaczowski B, Kern AD, Holloway AK, Begun DJ (2010) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, **187**, 245–260.
- Kondrashov AS (2003) Accumulation of Dobzhansky–Muller incompatibilities within a spatially structured population. *Evolution*, **57**, 151–153.
- Kruuk LE, Baird SJ, Gale KS, Barton NH (1999a) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, **153**, 1959–1971.
- Kruuk LEB, Gilchrist JS, Barton NH (1999b) Hybrid dysfunction in fire-bellied toads (*Bombina*). *Evolution*, **53**, 1611–1616.
- Lachance J, True JR (2010) X-autosome incompatibilities in *Drosophila melanogaster*: tests of Haldane's rule and geographic patterns within species. *Evolution*, **64**, 3035–3046.
- Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y (2002) Geographic structure in the European flat oyster, *Ostrea edulis* L, as revealed by microsatellite polymorphism. *Journal of Heredity*, **93**, 331–351.
- Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait and trait in a subdivided population under selection. *Genetics*, **164**, 1205–1219.
- Leaché AD, Cole CJ (2007) Hybridization between multiple fence lizard lineages in an ecotone: locally discordant variation in mitochondrial DNA, chromosomes, and morphology. *Molecular Ecology*, **16**, 1035–1054.
- Lehmann T, Diabate A (2008) The molecular forms of *Anopheles gambiae*: A phenotypic perspective. *Infection, Genetics and Evolution*, **8**, 737–746.
- Lemaire C, Allegrucci G, Naciri M *et al.* (2000) Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (*Dicentrarchus labrax*)? *Molecular Ecology*, **9**, 457–467.
- Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature*, **400**, 861–864.
- Lewontin RC (2002) Directions in evolutionary biology. *Annual Review of Genetics*, **36**, 1–18.
- Limborg MT, Pedersen JS, Hemmer-Hansen J, Tomkiewicz J, Bekkevold D (2009) Genetic population structure of European sprat (*Sprattus sprattus* L.): differentiation across a steep environmental gradient in a small pelagic fish. *Marine Ecology Progress Series*, **379**, 213–224.
- Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. *Science*, **325**, 1095–1098.
- Lu G, Bernatchez L (1998) Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchill). *Proceedings of the Royal Society B: Biological Sciences*, **265**, 1025–1030.

- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.
- MacCallum CJ, Nurnberger B, Barton NH, Szymura JM (1998) Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution*, **52**, 227–239.
- Malaua T, Bethenod M-T, Bontemps A *et al.* (2005) Assortative mating in sympatric host races of the European Corn Borer. *Science*, **308**, 258–260.
- Mallet J (1989) The genetics of warning colour in peruvian hybrid zones of *Heliconius erato* and *H. melpomene*. *Proceedings of the Royal Society B: Biological Sciences*, **236**, 163–185.
- Mallet J, Barton N (1989) Inference from clines stabilized by frequency-dependent selection. *Genetics*, **122**, 967–976.
- Masly JP, Jones CD, Noor MA, Locke J, Orr HA (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science*, **313**, 1448–1450.
- Matute DR, Novak CJ, Coyne JA (2009) Temperature-based extrinsic reproductive isolation in two species of *Drosophila*. *Evolution*, **63**, 595–612.
- Matute DR, Coyne JA (2010) Intrinsic reproductive isolation between two sister species of *Drosophila*. *Evolution*, **64**, 903–920.
- Mavarez J, Audet C, Bernatchez L (2009) Major disruption of gene expression in hybrids between young sympatric anadromous and resident populations of brook charr (*Salvelinus fontinalis* Mitchell). *Journal of Evolutionary Biology*, **22**, 1708–1720.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetics Research*, **23**, 23–35.
- McDaniel SF, Willis JH, Shaw AJ (2007) A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. *Genetics*, **176**, 2489–2500.
- Michel AP, Sim S, Powell TH *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences, USA*, **107**, 9724–9729.
- Miller JR, Hawthorne D (2005) Durability of marker-quantitative trait loci haplotypes in structured populations. *Genetics*, **171**, 1353–1364.
- Miranda MBB, Innes DJ, Thompson RJ (2010) Incomplete reproductive isolation in the blue mussel (*Mytilus edulis* and *M. trossulus*) hybrid zone in the northwest Atlantic: role of gamete interactions and larval viability. *Biological Bulletin*, **218**, 266–281.
- Mitton JB (1997) *Selection in Natural Populations*. Oxford University Press, New York.
- Montooth KL, Meiklejohn CD, Abt DN, Rand DM (2010) Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of *Drosophila*. *Evolution*, **64**, 3364–3379.
- Moore WS, Koenig WD (1986) Comparative reproductive success of yellow-shafted, red-shafted, and hybrid flickers across a hybrid zone. *The Auk*, **103**, 42–51.
- Moore WS, Price JT (1993) Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 196–225. Oxford University Press, New York.
- Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology*, **15**, 4229–4242.
- Navarro A, Barton NH (2003) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, **57**, 447–459.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Gronkjaer P (2003) Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Molecular Ecology*, **12**, 1497–1508.
- Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.
- Nikula R, Strelkov P, Väinölä R (2007) Diversity and trans-arctic invasion history of mitochondrial lineages in the North Atlantic *Macoma balthica* complex (Bivalvia: Tellinidae). *Evolution*, **61**, 928–941.
- Nikula R, Strelkov P, Väinölä R (2008) A broad transition zone between an inner Baltic hybrid swarm and a pure North Sea subspecies of *Macoma balthica* (Mollusca, Bivalvia). *Molecular Ecology*, **17**, 1505–1522.
- Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.
- Noor MA, Feder JL (2006) Speciation genetics: evolving approaches. *Nature Reviews of Genetics*, **7**, 851–861.
- Nosil P, Crespi BJ, Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, **417**, 440–443.
- Nosil P, Sandoval CP, Crespi BJ (2006) The evolution of host preference in allopatric vs. parapatric populations of *Timema cristinae* walking-sticks. *Journal of Evolutionary Biology*, **19**, 929–942.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
- Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology*, **16**, 5156–5157.
- Orr HA (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, **139**, 1805–1813.
- Orr MR (1996) Life-history adaptation and reproductive isolation in a grasshopper hybrid zone. *Evolution*, **50**, 704–716.
- Orr HA, Presgraves DC (2000) Speciation by postzygotic isolation: forces, genes and molecules. *Bioessays*, **22**, 1085–1094.
- Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Current Opinion in Genetics & Development*, **14**, 675–679.
- Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, 146–158.
- Panova M, Hollander J, Johannesson K (2006) Site specific genomic divergence in parallel hybrid zones suggests nonallopatric evolution of reproductive barriers. *Molecular Ecology*, **15**, 4021–4031.

- Patarnello T, Volckaert FA, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, **16**, 4426–4444.
- Pelc RA, Warner RR, Gaines SD (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, **36**, 1881–1890.
- Pelozuelo L, Malosse C, Genestier G, Guenego H, Frerot B (2004) Host-plant specialization in pheromone strains of the European Corn Borer *Ostrinia nubilalis* in France. *Journal of Chemical Ecology*, **30**, 335–352.
- Petersen MF, Steffensen JF (2003) Preferred temperature of juvenile Atlantic cod *Gadus morhua* with different haemoglobin genotypes at normoxia and moderate hypoxia. *Journal of Experimental Biology*, **206**, 359–364.
- Phadnis N, Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science*, **323**, 376–379.
- Pialek J, Barton NH (1997) The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics*, **145**, 493–504.
- Pickles A, Grahame J (1999) Mate choice in divergent morphs of *Littorina saxatilis* (Oliv): speciation in action? *Animal Behaviour*, **58**, 181–184.
- Pigeon D, Chouinard A, Bernatchez L (1997) Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of Lake Whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution*, **51**, 196–205.
- Powell JR, Petrarca V, della Torre A, Caccone A, Coluzzi M (1999) Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia*, **41**, 101–114.
- Prentice HC, Loenn M, Lefkovich LP, Runyeon H (1995) Associations between allele frequencies in *Festuca ovina* and habitat variation in the alvar grasslands on the Baltic island of Oeland. *Journal of Ecology*, **83**, 391–402.
- Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nature Reviews of Genetics*, **11**, 175–180.
- Ralph P, Coop G (2010) Parallel adaptation: One or many waves of advance of an advantageous allele? *Genetics*, **186**, 647–668.
- Rand DM, Harrison RG (1989) Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution*, **43**, 432–449.
- 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.
- Raufaste N, Orth A, Belkhir K *et al.* (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, **84**, 593–616.
- Rawson PD, Hayhurst S, Vanscoyoc B (2001) Species composition of blue mussel populations in the northeastern Gulf of Maine. *Journal of Shell Research*, **20**, 31–38.
- Rawson PD, Slaughter C, Yund PO (2003) Patterns of gamete incompatibility between the blue mussels *Mytilus edulis* and *M. trossulus*. *Marine Biology*, **143**, 317–325.
- Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. In: *Evolutionary Biology* (eds Dobzhansky T, Hecht MK, Steere WC), pp. 321–428. Plenum Press, New York.
- Remold SK, Lenski RE (2001) Contribution of individual random mutations to genotype-by-environment interactions in *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA*, **98**, 11388–11393.
- Rice WR, Holland B (1997) The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behavioral Ecology and Sociobiology*, **41**, 1–10.
- Ridgway G, Nævdal G (2004) Genotypes of *Mytilus* from waters of different salinity around Bergen. *Norway Helgoland Marine Research*, **58**, 104–109.
- Rieseberg LH, Buerkle CA (2002) Genetic mapping in hybrid zones. *American Naturalist*, **159**, S36–S50.
- Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.
- Riginos C, Cunningham CW (2005) Local adaptation and species segregation in two mussel (*Mytilus edulis* X *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, **14**, 381–400.
- Riginos C, Sukhdeo K, Cunningham CW (2002) Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Molecular Biology and Evolution*, **19**, 347–351.
- Robertson A (1975) Gene frequency distributions as a test of selective neutrality. *Genetics*, **81**, 775–785.
- Rogers SM, Bernatchez L (2006) The genetic basis of intrinsic and extrinsic postzygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus Clupeaformis*). *Journal of Evolutionary Biology*, **19**, 1979–1994.
- 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.
- Rolán-Alvarez E, Johannesson K, Erlandsson J (1997) The maintenance of a cline in the marine snail *Littorina saxatilis*: the role of home site advantage and hybrid fitness. *Evolution*, **51**, 1838–1847.
- Rolán-Alvarez E, Carballo M, Galindo J *et al.* (2004) Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. *Molecular Ecology*, **13**, 3415–3424.
- Ross CL, Harrison RG (2002) A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. *Evolution*, **56**, 2296–2312.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
- Sadedin S, Hollander J, Panova M, Johannesson K, Gavrilets S (2009) Case studies and mathematical models of ecological speciation 3. Ecotype formation in a Swedish snail. *Molecular Ecology*, **18**, 4006–4023.
- Santiago E, Caballero A (2005) Variation after a selective sweep in a subdivided population. *Genetics*, **169**, 475–483.
- Sarvas TH, Fevolden SE (2005) Pantophysin (Pan I) locus divergence between inshore v. offshore and northern v. southern populations of Atlantic cod in the north-east Atlantic. *Journal of Fish Biology*, **67**, 444–469.
- Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences, USA*, **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.

- Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat- and life-stage-specific survivorship of MPI genotypes. *Evolution*, **55**, 1336–1344.
- Schmidt PS, Serrão EA, Pearson GA *et al.* (2008a) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, **89**, S91–S107.
- Schmidt PS, Zhu C-T, Das J *et al.* (2008b) An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, **105**, 16207–16211.
- Shaw DD, Wilkins P (1980) Chromosome differentiation, hybrid breakdown and the maintenance of a narrow hybrid zone in *Caledia*. *Chromosoma*, **80**, 1–31.
- Shaw DD, Marchant AD, Contreras N *et al.* (1993) Genomics and environmental determinants of a narrow hybrid zone: Cause or coincidence? In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 165–195. Oxford University Press, Oxford.
- Shikano T, Ramadevi J, Merila J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined Sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution*, **27**, 2775–2789.
- Simard F, Ayala D, Kamdem GC *et al.* (2009) Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecology*, **9**, 17.
- Slatkin M (1975) Gene flow and selection in a two-locus system. *Genetics*, **81**, 787–802.
- Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genetics Research*, **71**, 155–160.
- Smadja G, Catalan J, Ganem C (2004) Strong premating divergence in a unimodal hybrid zone between two subspecies of the house mouse. *Journal of Evolutionary Biology*, **17**, 165–176.
- Smadja C, Galindo J, Butlin R (2008) Hitching a lift on the road to speciation. *Molecular Ecology*, **17**, 4177–4180.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, **64**, 2489–2509.
- Strelkov P, Nikula R, Väinölä R (2007) *Macoma balthica* in the White and Barents Seas: properties of a widespread marine hybrid swarm (Mollusca: Bivalvia). *Molecular Evolutionary Biology*, **16**, 4110–4127.
- Swenson NG, Howard DJ (2004) Do suture zones exist? *Evolution*, **58**, 2391–2397.
- Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, **166**, 581–591.
- Szymura JM (1993) Analysis of hybrid zones with *Bombina*. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 261–289. Oxford University Press, New York.
- Thériault V, Garant D, Bernatchez L, Dodson JJ (2007) Heritability of life-history tactics and genetic correlation with body size in a natural population of brook charr (*Salvelinus fontinalis*). *Journal of Evolutionary Biology*, **20**, 2266–2277.
- Toro JE, Innes DJ, Thompson RJ (2004) Genetic variation among life-history stages of mussels in a *Mytilus edulis*–*M. trossulus* hybrid zone. *Marine Biology*, **45**, 713–725.
- Tosto DS, Bidau CJ (1991) Distribution of chromosome frequencies within a hybrid zone of *Dichroplus pratensis* (Melanoplinae, Acrididae). *Heredity*, **67**, 299–306.
- Touré YT, Petrarca V, Traoré SF *et al.* (1998) Distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parasitologia*, **40**, 477–511.
- Turner JRG (1967) Why does the genotype not congeal? *Evolution*, **21**, 645–656.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.
- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, **42**, 260–263.
- Udovic D (1980) Frequency-dependent selection, disruptive selection, and the evolution of reproductive isolation. *American Naturalist*, **116**, 621–641.
- Väinölä R, Hvilson MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations. *Biological Journal of the Linnean Society*, **43**, 127–148.
- Väinölä R, Strelkov P (2011) *Mytilus trossulus* in Northern Europe. *Marine Biology*, DOI: 10.1007/s00227-0010-01609-z.
- Vasemägi A, Nilsson J, Primmer CR (2005) Expressed Sequence Tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution*, **22**, 1067–1076.
- 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 (1999) Reproductive isolation between sympatric races of pea aphids. I. gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.
- Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences, USA*, **106**(Suppl 1), 9939–9946.
- Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.
- Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, **54**, 1626–1637.
- Vines TH, Koehler SC, Thiel M *et al.* (2003) The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution*, **57**, 1876–1888.
- Waser NM (1993) Population structure, optimal outbreeding, and assortative mating in angiosperms. In: *The Natural History of Inbreeding and Outbreeding* (ed. Thornhill NW), pp. 173–199. University of Chicago Press, Chicago.
- Wheat CW, Watt WB, Pollock DD, Schulte PM (2006) From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* Phosphoglucose Isomerase (PGI). *Molecular Biology and Evolution*, **23**, 499–512.
- White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, **19**, 925–939.

- Wiebe KL (2000) Assortative mating by color in a population of hybrid Northern Flickers. *The Auk*, **117**, 525–529.
- 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.
- Wilhelm R, Hilbish TJ (1998) Assessment of natural selection in a hybrid population of mussels: Evaluation of exogenous vs endogenous selection models. *Marine Biology*, **131**, 505–514.
- Wolf JB, Lindell J, Backström N (2010) Speciation genetics: current status and evolving approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1717–1733.
- Wondji C, Simard F, Fontenille D (2002) Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Molecular and Biology*, **11**, 11–19.
- 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.

The authors are interested by diverse areas of population genetics, evolutionary biology and molecular evolution. N.B. and F.B. are members of the Institut des Sciences de l'Évolution at the University of Montpellier. They often focus their research on marine species and conduct their experiments at the marine lab in Sète. P.D. is member of the Centre d'Écologie Fonctionnelle et Evolutive in Montpellier and has long collaborated with N.B. and F.B., especially on speciation and population structure in marine bivalves. J.W. and E.L. were postdoctoral fellows of the Hi-Flo project coordinated by N.B. which aimed to broaden and improve the interpretations of genomic regions of enhanced differentiation. J.W. is now a lecturer in the Department of Genetics at the University of Cambridge and E.L. is post-doc at the Station biologique de Roscoff.

Data accessibility

Windows executables for simulations has been deposited at Dryad: doi:10.5061/dryad.8743

Supporting information

Additional supporting information may be found in the online version of this article.

Supporting File S1. Seven Windows platform executable files. Each file name contains a brief description of the simulation parameters and the number of the figure that presents one simulation outcome produced with the executable. To launch a simulation, open a file and to click on the "run" button.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the

authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

GLOSSARY

Genetic barrier to gene flow – A reduction in effective gene flow at neutral loci as a consequence of selection on linked loci. This reduction occurs because a neutral allele introduced from one gene pool into another, will be initially associated with alleles that are locally eliminated by either endogenous or exogenous selection. Recombination is thus necessary for the neutral immigrant allele to diffuse between the two gene pools. The diffusion of neutral alleles across such barriers is slower than over similar distances within a single gene pool.

Endogenous loci – Loci that produce hybrid fitness depression irrespective of the environment.

Endogenous barrier – A genetic barrier produced by endogenous loci.

Tension zone – Geographical zone where populations with incompatible genetic backgrounds are in contact (i.e. in which a cline of allele frequencies at endogenous loci can be observed).

Endogenous cline – A monotonic change in frequency of alleles at endogenous loci along some direction in space.

Exogenous loci – Loci at which different alleles are adapted to different environmental conditions.

Exogenous barrier – A genetic barrier produced by exogenous loci.

Exogenous cline – A monotonic change in frequency of alleles at exogenous loci along some direction in space.

Ecotone – The frontier, abrupt or gradual, between two different habitats.

Natural barrier to dispersal – a natural obstacle (mountain, river, unsuitable habitat, etc.) that locally reduces the probability of successful migration of individuals.

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.

Coarse-grained environment – A type of environmental heterogeneity in which dispersal is sufficiently low relative to the scale of the environmental variation for local adaptation to be easily maintained.

Strength of the barrier to gene flow – The decrease in the effective rate of migration.