

INVITED REVIEW

Divergent selection and heterogeneous genomic divergence

PATRIK NOSIL,*§ DANIEL J. FUNK† and DANIEL ORTIZ-BARRIENTOS‡

*Zoology Department and Biodiversity Research Centre, University of British Columbia, Vancouver BC, Canada, V6T 1Z4,
 †Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235, USA, ‡School of Integrative Biology,
 University of Queensland, St Lucia, Qld 4702, Australia

Abstract

Levels of genetic differentiation between populations can be highly variable across the genome, with divergent selection contributing to such heterogeneous genomic divergence. For example, loci under divergent selection and those tightly physically linked to them may exhibit stronger differentiation than neutral regions with weak or no linkage to such loci. Divergent selection can also increase genome-wide neutral differentiation by reducing gene flow (e.g. by causing ecological speciation), thus promoting divergence via the stochastic effects of genetic drift. These consequences of divergent selection are being reported in recently accumulating studies that identify: (i) 'outlier loci' with higher levels of divergence than expected under neutrality, and (ii) a positive association between the degree of adaptive phenotypic divergence and levels of molecular genetic differentiation across population pairs ['isolation by adaptation' (IBA)]. The latter pattern arises because as adaptive divergence increases, gene flow is reduced (thereby promoting drift) and genetic hitchhiking increased. Here, we review and integrate these previously disconnected concepts and literatures. We find that studies generally report 5–10% of loci to be outliers. These selected regions were often dispersed across the genome, commonly exhibited replicated divergence across different population pairs, and could sometimes be associated with specific ecological variables. IBA was not infrequently observed, even at neutral loci putatively unlinked to those under divergent selection. Overall, we conclude that divergent selection makes diverse contributions to heterogeneous genomic divergence. Nonetheless, the number, size, and distribution of genomic regions affected by selection varied substantially among studies, leading us to discuss the potential role of divergent selection in the growth of regions of differentiation (i.e. genomic islands of divergence), a topic in need of future investigation.

Keywords: chromosomal inversions, ecological genetics, F_{ST} , introgression, neutral gene flow, outlier loci, population genomics, QTL mapping, speciation

Received 21 February 2008; revision received 1 August 2008; accepted 27 August 2008

Introduction

Levels of genetic differentiation can be highly variable across the genome, a pattern we refer to here as 'heterogeneous genomic divergence' (Harrison 1991; Avise 2000; Rieseberg 2001; Via 2001; Wu 2001; Ortiz-Barrientos *et al.* 2002; Orr *et al.* 2004; Wu & Ting 2004; Gavrillets & Vose

2005; Mallet 2005; Turner *et al.* 2005; Harr 2006; Noor & Feder 2006; Begun *et al.* 2007; Mallet *et al.* 2007; Via & West 2008; see glossary for terminology). Genomic divergence may be particularly heterogeneous during the process of population divergence and speciation, during which genetic differentiation accumulates in some regions, while the homogenizing effects of gene flow or inadequate time for random differentiation by genetic drift precludes divergence in other regions (Wu 2001; Gavrillets & Vose 2005). Many factors potentially contribute to heterogeneous genomic divergence, including selection arising from ecological causes (Schluter 2000; Wu 2001) or genetic conflict (Rice 1998;

Correspondence: Patrik Nosil, Fax: +493089001-300.

E-mail: patrik.nosil@wiko-berlin.de

§Present address: Wissenschaftskolleg zu Berlin, Institute for Advanced study, Wallostrasse 19, 14193, Berlin.

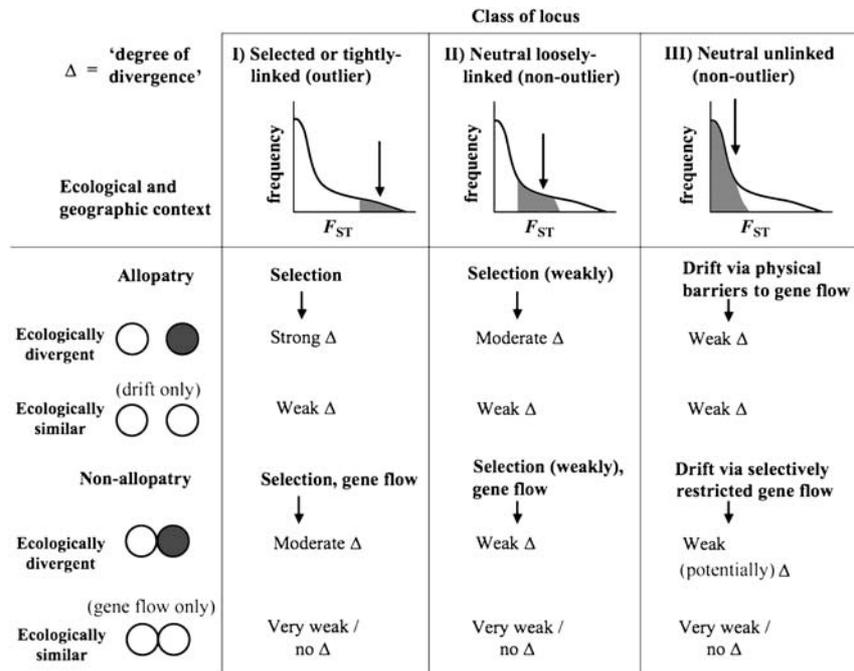


Fig. 1 Predicted patterns of genetic differentiation for different classes of loci (highlighted by grey coloration in the F_{ST} distributions) when divergent selection is present and absent (ecologically divergent and ecologically similar population pairs, respectively), for divergence both without and with gene flow (allopatry and nonallopatry, respectively). When divergent selection occurs, the processes expected to most strongly affect differentiation for each class of locus are noted above the expected degree of divergence. When divergent selection is absent, only drift or gene flow affect each class of locus (F_{ST} distributions would look similar but with a less extreme right-hand tail). Ecological divergence yields outlier status under both geographic scenarios. Class II loci should exhibit IBA in both allopatry and nonallopatry, while class III loci should only exhibit IBA under the nonallopatric scenario because neutral divergence in allopatry is independent of degree of ecological divergence. Patterns of heterogeneous genomic divergence are evident, such that genetic divergence generally varies among classes of loci as follows: class I > class II > class III.

Presgraves *et al.* 2003; Haig 2004; Arnqvist & Rowe 2005; Crespi 2007; Presgraves 2007), the stochastic effects of genetic drift (Kimura 1968, 1986; King & Jukes 1969; Ohta 1992, 2002), variable mutation rates (Balloux & Lugon-Moulin 2002; Hedrick 2005; Noor & Feder 2006), the genomic distribution and effect size of genes under selection (Orr 2005), and chromosomal structure (Noor *et al.* 2001; Rieseberg 2001; Ortiz-Barrientos *et al.* 2002).

We focus here on the contributions of divergent selection, defined as selection that acts in contrasting directions in two populations (cf. Schluter 2000; Rundle & Nosil 2005). Divergent selection itself can promote molecular genetic differentiation via two main mechanisms: (i) by acting on specific loci and those physically linked to them (Fisher 1930; Haldane 1930, 1932; Endler 1973; Lewontin & Krakauer 1973; Barton 2000), and (ii) by promoting reproductive isolation that causes barriers to gene flow (i.e. 'ecological speciation', Mayr 1963; Funk 1998; Schluter 2000; Rundle & Nosil 2005), thereby facilitating even genome-wide neutral divergence via genetic drift. The first mechanism involves a relatively direct role for selection in genetic differentiation and promotes divergence both in the presence and absence of gene flow (Fig. 1). The second

mechanism facilitates differentiation by a different process (drift). This second mechanism applies only to divergence with gene flow because in allopatric scenarios, divergent selection is not required to counter gene flow in order for neutral population differentiation to proceed. While these basic mechanisms account for heterogeneity in the origin and frequency of highly differentiated genomic regions, and in their degree of differentiation, selection may also affect the actual size of such 'islands of genomic divergence' on a chromosome (Turner *et al.* 2005; Harr 2006; Begun *et al.* 2007; Turner & Hahn 2007; see glossary).

Here, we review these roles for divergent selection in generating heterogeneous genomic divergence, and further consider the nature and growth of islands of genomic divergence. We consider mainly conceptual issues and empirical patterns, because methodology has been well covered elsewhere (e.g. Beaumont & Nichols 1996; Andolfatto 2001; Black *et al.* 2001; Schlötterer 2002; Luikart *et al.* 2003; Beaumont & Balding 2004; Beaumont 2005; Nielsen 2005; Storz 2005; Vasemagi & Primmer 2005; Hahn 2006; Hedrick 2006; Noor & Feder 2006; Bonin *et al.* 2007; Foll & Gaggiotti 2008; Riebler *et al.* 2008; Stinchcombe & Hoekstra 2008). We focus on divergence during the process

of population differentiation and speciation, and note that during this process, loci under divergent selection and loci causing reproductive isolation behave similarly, differentiating more strongly (even during allopatric divergence), and introgressing less freely than other loci (Barton 1979, 1983; Barton & Hewitt 1989; Mallet 1995, 2005, 2006; Wu 2001; Wu & Ting 2004; Nosil *et al.* 2005). While acknowledging this similarity (see the Supporting information for further discussion), we focus on divergent selection per se. Our frequent use of 'selection' is shorthand for 'divergent selection', while 'genetic differentiation' refers to 'molecular genetic differentiation'.

In the order presented, the specific aims of this study are to: (i) discuss theory and make explicit predictions about divergent selection and heterogeneous genomic divergence (Table 1), aided by the metaphor of genomic islands of divergence, (ii) review empirical studies testing for loci whose genetic divergence exceed neutral expectations, that is, 'outlier loci', which putatively represent the genetic signature of divergent selection, (iii) review empirical studies testing whether adaptive phenotypic divergence facilitates molecular genetic differentiation, (iv) describe how selection may promote the growth of genomic islands of divergence, and (v) integrate our findings and offer suggestions for future research.

How selection generates heterogeneous genomic divergence: processes and patterns

We classify loci into three categories according to how they are affected by selection: (i) loci under divergent selection and those tightly linked to them, (ii) loci loosely linked to those under selection, and (iii) neutral loci that are unlinked to selected loci. This categorization scheme is somewhat arbitrary because the effects of selection vary continuously, but is nonetheless useful because it relates to what can be empirically determined, for example, whether an affected locus is an outlier (see Fig. 1 and below).

Direct selection and its effect on tightly linked neutral loci (outlier loci)

Divergent selection on a locus will cause its adaptive divergence to a degree that often reflects a balance between the strength of selection and rates of gene flow (Fisher 1930; Haldane 1930, 1932; Wright 1931, 1940; Bulmer 1972; Endler 1973; Felsenstein 1976, 1981; Barton 1983; Slatkin 1985; Hendry *et al.* 2001; Butlin 2005). However, selection on one locus can also strongly affect the frequency of alleles at physically close or 'tightly linked' loci, even when the latter are selectively neutral. This process, whereby the frequency of alleles changes in concert with linked selected ones, has been termed genetic hitchhiking (Maynard Smith & Haigh 1974; Kaplan *et al.* 1989; for reviews see Barton

2000; Andolfatto 2001). In essence, under hitchhiking, the effect of selection on loci affecting fitness spills over to neutral loci. The effect of hitchhiking depends on a number of factors, but strongly on the ratio of recombination frequency and selection strength (r/s), with stronger selection and lower recombination yielding greater effects on linked loci (Bodmer & Parsons 1962; Felsenstein 1981; Charlesworth *et al.* 1997; Barton 2000; Andolfatto 2001; Ortiz-Barrientos *et al.* 2002; Butlin 2005). This concept is illustrated in Fig. 2A and also implicit in Fig. 2B, which further illustrates how genetic divergence at neutral loci decreases with distance from a selected locus. Hitchhiking can also have stochastic effects; by reducing N_e at loci affected by selection, it can amplify divergence via genetic drift (Robertson 1961; Hill & Robertson 1966; Santiago & Caballero 1998).

Thus, loci under selection and those tightly physically linked to them should exhibit greater differentiation than distantly linked or unlinked neutral regions (Fig. 2; Lewontin & Krakauer 1973; Bowcock *et al.* 1991; Beaumont & Nichols 1996; Black *et al.* 2001; Schlötterer 2002; Luikart *et al.* 2003; Beaumont & Balding 2004; Beaumont 2005; Nielsen 2005; Storz 2005; Vasemagi & Primmer 2005; Nachman 2006; Foll & Gaggiotti 2008; Riebler *et al.* 2008; Stinchcombe & Hoekstra 2008). Divergent selection thus results in outlier loci whose genetic divergence exceeds neutral expectations. When testing for outliers, simulations are generally used to determine the upper level of genetic divergence expected under neutrality, and loci whose genetic differentiation exceeds this neutrality threshold are deemed outliers (Fig. 3A). Genetic differentiation itself is often quantified using F_{ST} , a fixation index that measures the degree of population differentiation (Wright 1940; Hudson *et al.* 1992). A final point is that new mutations are more likely to diverge between populations if they arise in genomic regions already under divergent selection (i.e. that already exhibit reduced introgression). Consequently, regions of strong differentiation are predicted to accumulate in clusters within the genome, rather than being randomly distributed throughout it (see discussion of genomic island growth) (Rieseberg 2001; Navarro & Barton 2003; Gavrillets 2004, chapter 6; Kirkpatrick & Barton 2006).

Here, we review findings from studies testing for outlier loci in natural populations. Other methods of evaluating the molecular signatures of selection exist, such as McDonald-Kreitman tests and candidate gene approaches (see Nielsen 2005; Nachman 2006 for reviews; O'Malley *et al.* 2007). However, because our study concerns population-level patterns of genomic heterogeneity, we focus on methods that can evaluate divergent selection in large numbers of gene regions across the genome. In particular, we discuss 'genome scans', in which many individuals are screened for variation in dozens or hundreds of molecular markers (Lewontin & Krakauer 1973; Bowcock *et al.* 1991; Beaumont

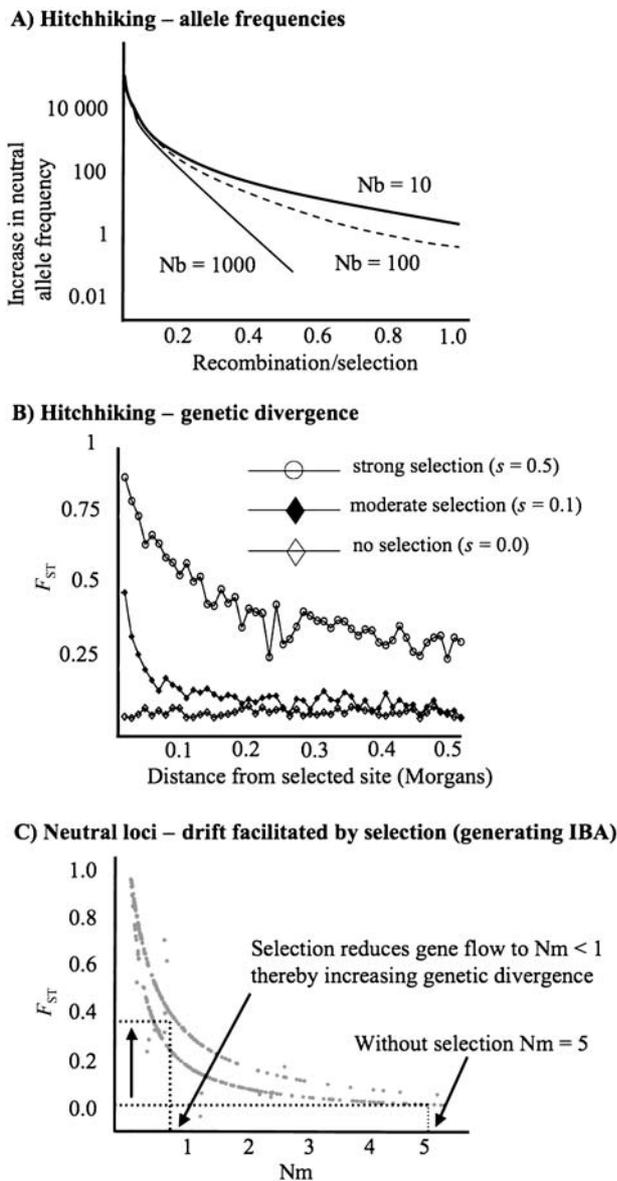


Fig. 2 Divergent selection's contributions to heterogeneous genomic divergence. (A) Genetic hitchhiking – allele frequencies. The net increase in allele frequency of a neutral allele (y -axis) via hitchhiking, under the spread of an advantageous new mutation (Barton 2000 for further details on scaling of the axes). N_b refers to neighbourhood (i.e. population) size. Modified from Barton (2000) and reprinted with permission of the Royal Society of London. (B) Hitchhiking – genetic divergence. Simulation results showing the effects of divergent selection on levels of differentiation at linked neutral sites, for two demes exchanging migrants (background selection present, but similar patterns observed in its absence). Modified from Charlesworth *et al.* (1997, see for details) with permission of Cambridge University Press. (C) Relationship between neutral F_{ST} and N_m across empirical studies of animals (data from Morjan & Rieseberg 2004). Note that F_{ST} increases when divergent selection reduces N_m (e.g. by selecting against immigrants). For example, reducing N_m from five to just below one considerably increases F_{ST} (see thick arrow).

& Nichols 1996; Andolfatto 2001; Black *et al.* 2001; Vitalis *et al.* 2001, 2003; Schlötterer 2002; Luikart *et al.* 2003; Beaumont & Balding 2004; Beaumont 2005; Nielsen 2005; Storz 2005; Vasemagi & Primmer 2005; Hahn 2006; Hedrick 2006; Noor & Feder 2006; Bonin *et al.* 2007; Foll & Gaggiotti 2008; Riebler *et al.* 2008; Stinchcombe & Hoekstra 2008). By distinguishing outliers from putatively neutrally evolving (nonoutlier) loci, genome scans enable the quantification of regions under selection, the evaluation of their distribution across the genome [e.g. by evaluating linkage disequilibrium (LD)], tests for replicated differentiation across population comparisons, the comparison of evolutionary patterns at outlier vs. neutral loci, etc. Genome scans therefore allow the investigation of how and why divergence varies across the genome, outstanding questions in evolutionary genetics (Orr 2005). Genome scans typically employ anonymous molecular markers whose chromosomal positions are unknown, but may also be combined with classical genetic approaches such as quantitative trait locus (QTL) mapping (Rogers & Bernatchez 2005; Via & West 2008; see Fig. 3B.)

Loosely linked neutral loci: additional effects of hitchhiking

Loci that are too far removed along a chromosome from a selected site for hitchhiking to elevate them to outlier status may nonetheless be somewhat influenced by selection, because hitchhiking effects can extend a considerable distance from the selected locus (Charlesworth *et al.* 1997; Nielsen 2005; Via & West 2008; Figs. 2B and 3B). Such nonoutlier loci are referred to as loosely linked (= weakly linked) and are expected to exhibit moderately elevated differentiation compared to completely unlinked neutral loci. Selection acting on loosely linked loci thus further contributes to heterogeneous genomic differentiation (Fig. 1). These effects of loose linkage were characterized by Charlesworth *et al.* (1997) using a combination of simulations and analytical theory. These authors found that strong divergent selection ($s = 0.5$) resulted in elevated differentiation of neutral loci relative to the case where divergent selection was absent, with such effects extending far along the chromosome and persisting despite high recombination rates. Under moderate selection ($s = 0.1$), differentiation was less elevated from neutral expectations and approached neutral expectations when the distance from the selected site exceeded two Morgans (Fig. 2B). Thus, the effects of hitchhiking on genetic differentiation are positively related to the strength of selection.

These observations yield the prediction that at neutral loci loosely linked to those under selection, levels of genetic differentiation among population pairs will be positively correlated with degree of adaptive phenotypic divergence (a proxy for the strength of divergent selection) across population pairs (Fig. 3C). We hereafter refer to such associations as 'isolation by adaptation' (IBA, following Nosil

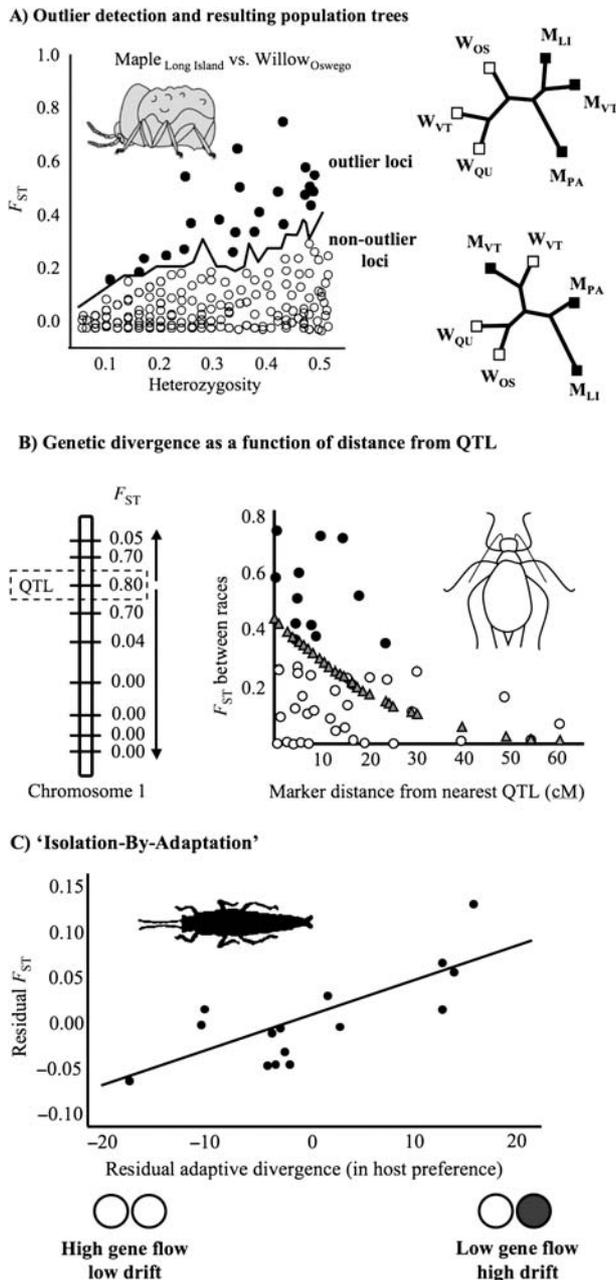


Fig. 3 Empirical patterns illustrating heterogeneous genomic divergence. (A) Outlier detection using the genome scan approach, for a pair of *Neochlamisus bebbianae* leaf beetle populations. Simulations determine the upper level of genetic divergence expected under neutrality, and loci that exceed this 'neutrality threshold' (solid line, in this case 95% quantile) are inferred to have evolved under divergent selection. If outliers are highly replicated across population pairs that have diverged ecologically, then population trees from pooled outlier loci are likely to group populations according to ecology (e.g. box colour indicates host plant), reflecting divergent selection. In contrast, trees from putatively neutral nonoutlier loci may group populations according to geography, reflecting spatial opportunities for gene flow. Modified from Egan *et al.* (2008) and reprinted with permission of the Society for the Study of Evolution. (B) Genetic

differentiation between clover- and alfalfa-associated pea aphids as a function of distance from QTL. On the left, the panel depicts a hypothetical decrease in F_{ST} along a chromosome with increasing distance from a QTL. On the right, real data depict markers that were outliers (filled circles) and nonoutliers (open circles). Triangles show the predicted values from a logistic regression of outlier status on distance to the nearest QTL. Modified from Via & West (2008) and reprinted with permission of Blackwell Publishing. (C) Pattern depicting IBA, in which neutral genetic differentiation between population pairs of *Timema cristinae* walking-stick insects is positively correlated with their degree of adaptive divergence (here measured as the degree of divergence in host plant preferences), independent of geographic distance. Modified from Nosil *et al.* (2008) and reprinted with permission of the Society for the Study of Evolution.

Unlinked neutral loci: general barriers to neutral gene flow

Divergent selection can also have more indirect yet widespread effects on heterogeneous genomic divergence, by reducing gene flow between populations (Barton & Bengtsson 1986; Pialek & Barton 1997; Gavrillets & Cruzan 1998; Gavrillets 2004) and thereby facilitating neutral divergence across the genome via genetic drift (Fig. 2C). The basic scenario is one in which a population is subject to continuous immigration. Due to divergent local adaptation, immigrants have lower fitness than residents, yielding selection against immigrants (Mallet & Barton 1989; Funk 1998; Via *et al.* 2000; Hendry 2004; Nosil *et al.* 2005 for review). In this fashion, selection against incoming, locally maladapted alleles will act as a 'general barrier' to the spread of neutral alleles between populations (Bengtsson 1985; Barton & Bengtsson 1986; Pialek & Barton 1997; Gavrillets & Cruzan 1998; Navarro & Barton 2003; Gavrillets & Vose 2005).

The effective immigration rate of neutral alleles is slowed even further under assortative mating (Gavrillets 2004, p. 148; for a theoretical summary, see the Supporting information). For example, when populations evolve preferences for their native habitat, the resulting decrease in between-habitat dispersal may reduce opportunities for between-population mating, yielding habitat isolation and reducing gene flow between populations. In fact, any reproductive barrier, by definition, reduces gene flow,

thereby potentially facilitating neutral genetic divergence via drift. Notably, adaptive divergence can promote the evolution of all such barriers, including 'nonecological' ones such as intrinsic hybrid inviability (Bateson 1909; Dobzhansky 1936, 1937, 1951; Muller 1940, 1942; Orr 1995; Orr & Turelli 2001; Gavrillets 2004; Dettman *et al.* 2007), via by-product models of ecological speciation (Mayr 1947, 1963; Funk 1998; Lu & Bernatchez 1999; Schluter 2000; Ogden & Thorpe 2002; Rundle & Nosil 2005; Funk *et al.* 2006; Vines & Schluter 2006). We hereafter refer to this scenario, whereby selection facilitates drift by reducing gene flow, as the 'general barriers' mechanism.

The 'general barriers' mechanism can generate heterogeneous genomic divergence due the stochastic nature of drift, which causes different neutral loci within a population to differentiate to varying degrees. The process also predicts that more adaptively divergent populations will experience greater gene flow reduction and associated neutral differentiation, yielding the pattern of IBA (Thibert-Plant & Hendry, 2008). Under the general barriers scenario, IBA is particularly analogous to IBD, as both can affect neutral loci unlinked to those under selection. An open empirical question is how commonly adaptive divergence restricts gene flow strongly enough to generate IBA at such loci, given that even small amounts of gene flow can overwhelm the ability of drift to cause neutral differentiation (Fig. 2C; Wright 1931, 1940; Barton & Bengtsson 1986).

Another issue is that if time since population divergence is the main predictor of adaptive and genetic divergence, then the pattern of IBA may arise without general barriers, simply because both types of divergence increase with time (Rogers 1986; Lande 1992; Whitlock 1999; Merila & Crnokrak 2001; McKay & Latta 2002). Nonetheless, theory indicates that IBA may commonly be generated by general barriers because gene flow represents the predominant force affecting levels of genetic differentiation, even under low migration rates (Wright 1931, 1943; Slatkin 1993; Beaumont & Nichols 1996; Hartl & Clark 1997; Rousset 1997; Balloux & Lugon-Moulin 2002; Hedrick 2005). From an empirical perspective, recently developed coalescent-based methods can be used to estimate migration rate (m) separately from divergence time (Nielsen & Wakeley 2001; Hey & Nielsen 2004; Hey 2006; Nosil 2008), thereby allowing IBA to be tested while controlling for time since divergence. Another point is that if IBA arises via selective processes, then absolute values of adaptive divergence (e.g. quantitative trait divergence measured using Q_{ST}) are expected to exceed those of F_{ST} , despite the correlation of these two measures (Rogers 1986; Lande 1992; Whitlock 1999). Thus, time vs. adaptation can potentially be distinguished as alternative causes of IBA.

In our review, we focus on the exhibition of IBA by neutral loci that are not tightly linked to those under selection, because such IBA is an expression of the somewhat coun-

terintuitive effects of selection on nonselected genomic regions (via either loose linkage or general barriers). We discuss factors affecting the prevalence and extent of IBA. For example, under the general barriers scenario, the proportion of loci exhibiting IBA should vary according to factors that affect genetic drift, such as levels of gene flow and effective population size. Another prediction is that if IBA is detected in the absence of gene flow (e.g. at spatial scales beyond that at which gene flow occurs), then it is more likely to have arisen via loose linkage than via general barriers.

Genomic islands of divergence: an integrated view of genomic heterogeneity

To help understand how the above processes combine to generate heterogeneous genomic divergence, we use the concept of 'genomic islands of divergence' (Wu 2001; Turner *et al.* 2005; Harr 2006; Turner & Hahn 2007). Here, we define a 'genomic island' as any gene region, be it a single nucleotide or an entire chromosome, that exhibits significantly greater differentiation than expected under neutrality. We conceptually extend the metaphor of genomic islands of divergence by comparing the heterogeneous genetic differentiation observed along a chromosome to the topography of oceanic islands and the contiguous sea floor to which they are connected. The concept is visually depicted in Fig. 4.

Following this metaphor, sea level represents the threshold above which observed differentiation is significantly greater than expected by neutral evolution alone. Thus, an island is composed of loci — both selected (dark grey) and tightly linked neutral (white) loci — that should be identifiable as outliers in a genome scan. Island elevation (differentiation) is a function of selection strength. Selection also contributes to island size, representing the length of contiguous highly differentiated chromosome. Loosely linked (light grey) loci are depicted as regions far enough from selected loci to fall below sea level as nonoutliers, but still close to the surface, being more highly differentiated than most unlinked neutral loci. The differentiation of such loci will also be a function of selection strength, yielding IBA. Farther still from the selected locus, the sea floor drops more steeply, with variation in its topography representing the variable levels of baseline neutral differentiation occurring by genetic drift alone. Thus, the sea floor may be rather homogeneous during divergence with gene flow, where gene exchange between populations homogenizes most neutral differentiation. However, if selection promotes general barriers to gene flow during such nonallopatric divergence, genome-wide differentiation under drift is facilitated. This is illustrated by a shallow sea with undersea mounts whose proximity to sea level will vary as a function of selection strength (light grey caps depict loci whose

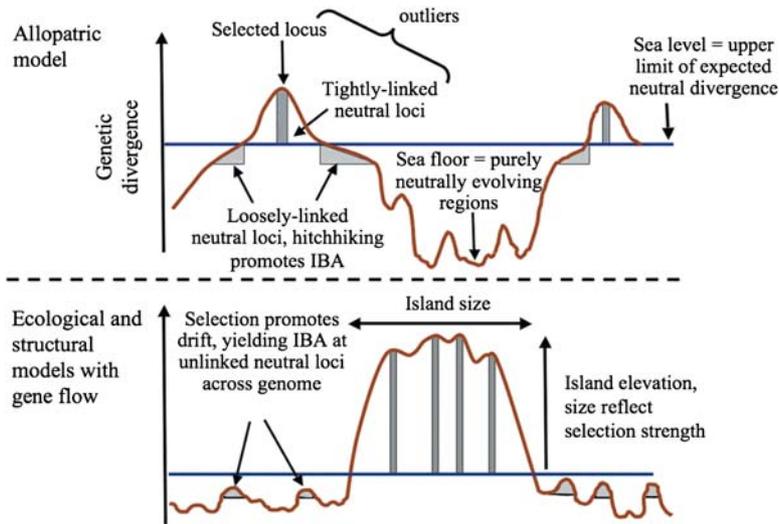
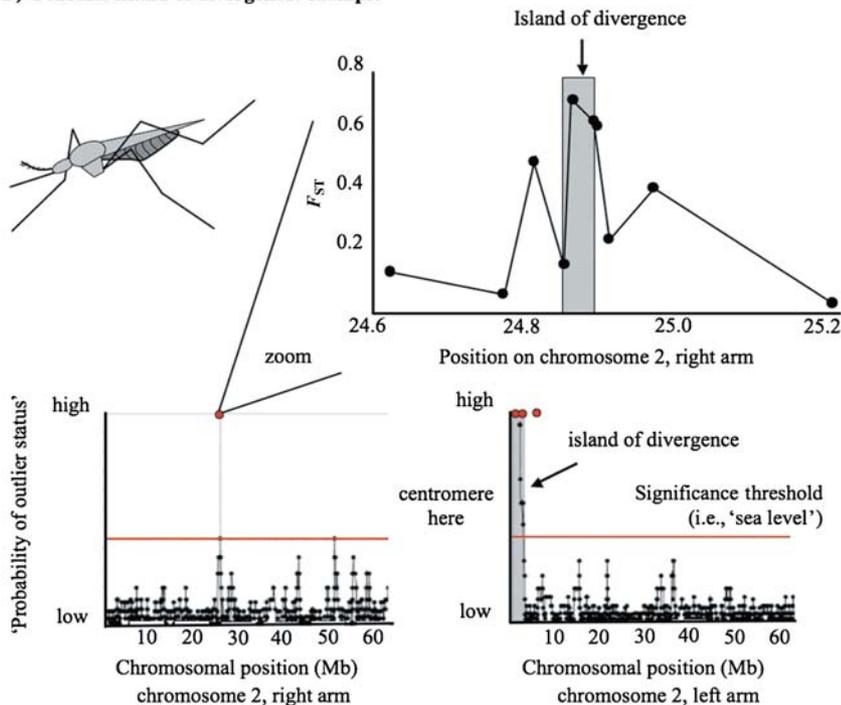
A) Genomic island of divergence: visual metaphor**B) Genomic island of divergence: example**

Fig. 4 The metaphor of genomic islands of divergence. (A) Schematic illustration of expected patterns of differentiation along a chromosome under models with and without gene flow. Sea level represents a 'neutrality threshold'. Islands are genomic regions exhibiting greater differentiation than expected under neutrality, thereby rising above sea level. See text for further details. (B) An empirical example involving incipient species of *Anopheles gambiae*. The bottom two panels depict patterns of differentiation across chromosome 2 (Turner *et al.* 2005). Grey areas were identified as highly differentiated in sliding window analyses, with differentiation further confirmed by sequencing loci within these regions (red circles). A large island is evident on the left arm, near the centromere. A small island is also evident on the right arm. The top panel treats a subsequent study (Turner & Hahn 2007) where portions of all annotated genes within the smaller island were sequenced. As predicted by Turner *et al.* (2005), sequence differentiation peaked within the 'island' (between the grey lines). However, the fine-scale data from Turner & Hahn (2007) allow more detailed characterization of the nature of the island, showing, for example, that differentiation drops off rapidly with distance from the region of maximum differentiation (i.e. the island is very steep). Modified from the original studies and reprinted with permission of the Public Library of Science and the Society of Molecular Biology and Evolution.

differentiation was facilitated by general barriers such that they contribute to IBA). In contrast, the sea floor is expected to be more heterogeneous during allopatric divergence, where the absence of gene flow allows the stochasticity of divergence due to drift and variable mutation rates to be more fully expressed (Balloux & Lugon-Moulin 2002; Hedrick 2005). Notably, the islands themselves may grow through time, because new mutations are more likely to differentiate if they arise in or adjacent to regions already subject to divergent selection, thus contributing to island

growth (Rieseberg 2001; Navarro & Barton 2003; Gavrilets 2004, chapter 6; Kirkpatrick & Barton 2006).

Literature reviews

We conducted two literature reviews to evaluate critical issues relating to genomic heterogeneity. First, we review genome scan studies seeking to identify outlier loci exhibiting elevated divergence (Table 2). Second, we review studies that allow the evaluation of IBA (Table 3).

Table 1 Predictions concerning the roles of divergent selection role in heterogeneous genomic divergence. Both effects on particular gene regions (see also Fig. 1) and models of the growth of differentiated regions (i.e. 'islands'; Fig. 4) are treated

| Type of locus | Predictions |
|---|--|
| Locus directly under selection | – strong genetic divergence at these loci (as illustrated by, e.g. outlier status) |
| Locus tightly physically linked to those under selection | – strong genetic divergence at these loci (as illustrated by, e.g. outlier status) |
| Locus loosely physically linked to those under selection | – moderately increased genetic divergence compared to neutrality – IBA pattern should be observed even beyond the spatial scale of gene flow (e.g. among completely allopatric populations) |
| Neutral, unlinked loci (affected by 'general barriers') | – IBA expected at the spatial scale of gene flow <i>if</i> gene flow is sufficiently reduced to allow divergence via genetic drift (i.e. selection and reproductive isolation must be strong); – likelihood of IBA increases with decreasing N_e (i.e. as drift becomes more effective) |
| Models for growth of differentiated regions | Predictions |
| I. Allopatric model | – many differentiated regions (i.e. 'genomic islands') – islands need not be clustered within the genome, and will often be small in size – number and elevation of islands increases with time since population divergence |
| II. Ecological model (divergent selection with gene flow) | – large islands will occur – genomic clustering of islands, with genes affecting local adaptation and reproductive isolation residing within these clusters, perhaps fewer islands than the allopatric model – islands need not involve chromosomal inversions (e.g. supergenes) – small islands, which have not yet grown, are also possible |
| III. Structural model | – genomic differentiation is facilitated by chromosomal inversions and other factors that reduce recombination, dependent on the degree to which genes affecting local adaptation and reproductive isolation reside within inversions – the extent of this facilitation also depends on how far outside an inversion the introgression reducing effects of the inversion extend – islands can be larger than in the other models, and persist for longer periods of time |

Literature review 1: genome scans and outlier loci

Numerous methodological issues arise when interpreting genome scans, including aspects of experimental design and data collection. Major issues include the relationship between type I error and outlier detection (reflecting the multiple comparisons involved in screening large numbers of loci), disentangling the effects of mutation rate variation, demography, and selection on levels of genetic divergence, and determining the ecological causes of the divergent selection. We report on the methodological robustness of each study described here in the Supporting information, and refer readers to the previous reviews cited in the Introduction for more detailed treatment of methodological issues.

Identifying relevant genome scan studies Relevant papers were identified from a Web of Science (WOS) search on 'genome scan and selection', and from the studies cited by these papers. In order to focus on genome-wide patterns in natural populations, our review excluded studies that were genomically restricted (e.g. those evaluating a single chromosome or assaying markers known a priori to be linked to QTL), treated humans or domesticated species, or did not use divergence-based methods (e.g. those evaluating selective sweeps within populations), although we cite

such studies where appropriate. This approach located 20 relevant studies (Table 2; a few QTL studies are treated separately), a modest number that demonstrates the nascent state of this field. Nonetheless, these studies illustrated a number of emerging patterns, which are discussed below.

Proportion of genome exhibiting outlier behaviour We found substantial, but not extreme, variation in the proportion of loci identified as outliers across studies. This observation holds for studies considering all reported outlier loci (range: 0.4–24.5%, mean: 8.5%, $n = 18$), those that only tallied outliers observed in multiple population pairs (i.e. studies examining 'replicated divergence': range: 1.4–12.0%, mean: 5.8%, $n = 7$), and studies examining the distribution of outliers within and between ecological types (range: 0.4–9.0%, mean: 4.5%, $n = 5$). Note that these percentages quantify the proportion of loci that were outliers in at least one (or, in the case of replicated outliers, at least two) of a study's pairwise population comparisons, rather than mean proportions across individual comparisons. These estimates should be interpreted with great caution, as different studies varied in the number of populations and individuals examined, molecular markers employed, methods for estimating baseline neutral differentiation and criteria determining outlier status (see Table 2 and the

Table 2 Summary of genome scan studies. Provided is the study system, a description of the divergent forms being compared, the type and number of markers used, the study design ('compare divergence' = compare population divergence among loci), the main result with respect to outliers, patterns observed in population trees, and levels of linkage disequilibrium (LD) among outlier loci. More specifically, the 'Population trees differ?' column addresses whether population trees from neutral (nonoutlier) loci group populations by geography, whereas trees that include (or only use) outlier loci group populations by ecology. Here, 'Yes' indicates that outlier-including trees result in groupings that are consistent with monophyly of the same ecological type, and 'Partially' indicates that outlier-including trees are more structured by ecology than neutral trees but do not provide evidence of complete monophyly by ecological type. For detailed consideration of the methodological robustness of each study, see the Supporting information

| Study system | Divergent forms | Data (marker, no. of loci) | Study design | Main result – outliers | Population trees differ? | LD? | Reference |
|--|--|--|---|---|--------------------------|---|---|
| Studies examining parallel divergence/association with ecological parameters | | | | | | | |
| 1. <i>Littorina saxatilis</i> (intertidal snails) | Upper vs. lower shore ecotypes | 306 AFLP loci | Compare divergence between sympatric ecotypes at three different shores | 5% of loci consistently outliers between ecotypes at all three shores | Yes | Low, except in middle of cline, implying any LD is generated by migration | Wilding <i>et al.</i> 2001; Grahame <i>et al.</i> 2006; Wood <i>et al.</i> 2008 |
| 2. <i>Coregonus clupeaformis</i> (whitefish) | Dwarf vs. normal lake ecotypes | 440 AFLP loci | Compare divergence between sympatric ecotypes from four different lakes | up to 3.2% of loci outliers between ecotypes, 1.4% outliers in all four lakes | Yes | Possibly moderate, as outlier loci were nonrandomly distributed among primer combinations | Campbell & Bernatchez 2004 |
| 3. <i>Salmo salar</i> (Atlantic salmon) | Saltwater vs. brackish vs. freshwater habitats | 95 genomic and EST-derived mini- and microsatellites | Compare divergence for four population pairs occupying different habitats, at local and global spatial scales | 23–26% of loci outliers, 9% of EST derived loci were outliers at different spatial scales, population pairs, and geographic regions | N/A | N/A | Vasemagi <i>et al.</i> 2005 |
| 4. <i>Rana temporaria</i> (common frog) | Altitudinal gradient | 392 AFLP loci | Compare divergence between twelve population pairs differing in altitude | 8–14% of loci outliers, roughly 2% of loci involved in altitudinal adaptation specifically | Yes | N/A | Bonin <i>et al.</i> 2006 |
| 5. <i>Hylobius abietis</i> (pine weevil) | N/A | 83 AFLP loci | Compare divergence among six forest regions, and relate genetic divergence to environmental variables | 6% of loci outliers, two loci (2.4%) strongly correlated with variation in environmental parameters | N/A | N/A | Joost <i>et al.</i> 2007 |

Table 2 Continued

| Study system | Divergent forms | Data (marker, no. of loci) | Study design | Main result—outliers | Population trees differ? | LD? | Reference |
|--|--|--|--|--|--------------------------|--|------------------------------|
| Studies examining divergence within vs. between ecological forms | | | | | | | |
| 6. <i>Zeiraphera diniana</i> (herbivorous insect) | Sympatric larch- vs. pine-feeding host forms | 1291 AFLP loci | Compare heterogeneity of genomic divergence between sympatric host forms (four comparisons) vs. between populations of the same form (six comparisons) | heterogeneity of genomic divergence between sympatric host forms, but not between geographic populations of the same form | N/A | Possibly high, reflecting concentration of loci involved in between-race differentiation on just a few chromosomes | Emelianov <i>et al.</i> 2004 |
| 7. <i>Cryptomeria japonica</i> (coniferous tree) | Slender branched, soft-leaved vs. rough branched hard-leaved variety | 139 CAPS markers based on cDNA sequences | Compare divergence using 29 populations, considering divergence among populations within and between varieties | 10% of loci outliers, 6.5% outliers only between varieties | N/A | N/A | Tsumura <i>et al.</i> 2007 |
| 8. <i>Timema cristinae</i> (walking-stick insect) | <i>Adenostoma</i> and <i>Ceanothus</i> host plant ecotypes | 534 AFLP loci | Compare divergence between nine population pairs on different hosts and between six population pairs on the same host | 8% of loci outliers in multiple comparisons, 1–2% outliers only in between-host comparisons with weak parallel divergence among pairs | Partially | Low, but higher than for nonoutliers within both allopatric and parapatric populations, consistent with some weak physical linkage | Nosil <i>et al.</i> 2008 |
| 9. <i>Neochlamisus bebbinae</i> (leaf beetle) | Willow and maple host forms | 447 AFLP loci | Compare divergence between nine population pairs on different hosts and between six population pairs on the same host | 15% of loci outliers in multiple comparisons, 5% outliers only in between-host comparisons with strong parallel divergence among pairs | Yes | Low | Egan <i>et al.</i> 2008 |

Table 2 *Continued*

| Study system | Divergent forms | Data (marker, no. of loci) | Study design | Main result – outliers | Population trees differ? | LD? | Reference |
|--|--|---|---|---|---|---|---|
| 10. <i>Phytomyza glabricola</i> (holly leafminer) | <i>Ilex glabra</i> vs. <i>I. coriacea</i> holly host forms | 45 AFLP loci | Compare divergence between two population pairs on different hosts and between two pairs on the same host | 18% of loci outliers, 11–16% of loci outliers in a between-host pair with 9% of loci outliers in both between-host pairs | N/A; but mtDNA groups by geography whereas AFLPs group by host (i.e. ecology) | N/A | Scheffer & Hawthorne 2007 |
| 11. <i>Diabrotica virgifera</i> (western corn rootworm) | Crop rotation resistant vs. wild type strains | 253 AFLP loci | Compare divergence for nine between-strain population pairs and between six pairs of the same strain | % overall outliers not reported, < 0.5% (one locus) an outlier in between-strain comparisons only | N/A | N/A | Miller <i>et al.</i> 2007 |
| 12. <i>Zostera marina</i> (marine flowering plant) | Tidal creek vs. tidal flat | 25 EST-derived and anonymous microsatellite markers | Compare divergence between populations in different habitats for three independent population pairs | 12% of loci outliers, with these exhibiting replicated divergence for population pairs in different habitats, but not being outliers in same-habitat population pairs | N/A | N/A | Oetjen & Reusch 2007 |
| Other types of studies | | | | | | | |
| 13. <i>Quercus robur</i> L. and <i>Q. petraea</i> (oaks) | High nutrient soil vs. drier habitat species | 389 markers (isozymes, AFLPs, SCARs, microsatellites, and SNPs) | Compare divergence between species using 7–10 population pairs, depending on the marker | 12% of loci outliers | N/A | Low, outlier loci distributed among many different linkage groups | Scotti-Saintagne <i>et al.</i> 2004 |
| 14. <i>Picea abies</i> (spruce) | N/A | 125 AFLPs, 25 SSRs and 2 EST mapped markers | Compare divergence for three pairs of populations | 6% of loci outliers, but not replicated across population pairs | N/A | Low, outlier loci distributed among many different linkage groups | Acheré <i>et al.</i> 2005 |
| 15. <i>Anopheles gambiae</i> (African malaria mosquito) | M vs. S form | Hybridized DNA of single mosquitoes from samples of M and S forms to microarray chips (1577 probes) | Compare divergence between forms using seven M samples and seven S samples | 1.2% of the genome, from only three genomic regions, highly differentiated between forms | N/A | N/A | Turner <i>et al.</i> 2005; Turner & Hahn 2007 |

Table 2 Continued

| Study system | Divergent forms | Data (marker, no. of loci) | Study design | Main result – outliers | Population trees differ? | LD? | Reference |
|---|--|--|--|--|--------------------------|-----|-------------------------------|
| 16. <i>Mus musculus</i> (house mouse) | <i>Mus musculus musculus</i> vs. <i>M. m. domesticus</i> | > 10 000 SNP markers | Compare divergence between subspecies using 22 wild-derived inbred strains | 7.5% of the autosomal genome highly differentiated, comprising eight genomic regions | N/A | N/A | Harr 2006 |
| 17. <i>Crassostrea virginica</i> (oyster) | N/A | 215 AFLP loci | Compare divergence between a population pair | 1.4% of loci outliers | N/A | Low | Murray & Hare 2006 |
| 18. <i>Fagus sylvatica</i> (beech tree) | Temperature gradient | 254 AFLP loci | Compare divergence between five population pairs differing in temperature | < 0.5% of loci (one locus) outliers, associated with divergence in temperature | N/A | N/A | Jump <i>et al.</i> 2006 |
| 19. <i>Howea forsteriana</i> and <i>H. belmoreana</i> (oceanic palms) | Divergent flowering times and soil preferences | 274 AFLP loci | Compare divergence between two species | 1% of loci outliers | N/A | N/A | Savolainen <i>et al.</i> 2006 |
| 20. <i>Peromyscus</i> spp. (mice) | N/A | 10–37 protein-coding genes, depending on mouse species | Compare divergence among populations within six different species | 10.5% outliers (percent outliers within data sets ranged from 0–30), parallel divergence for some loci | N/A | N/A | Storz & Nachman 2003 |

Table 3 Summary of studies providing information on isolation by adaptation (IBA), where IBA refers to a positive association between the degree of adaptive phenotypic (or ecological) divergence between populations and the level of genetic differentiation at putatively neutral markers. All findings of IBA provided here control for geographic distance. Detection of isolation by distance (IBD) is also noted. Three main types of studies are reported: (i) population genetic studies explicitly examining IBA, generally using distance matrices of adaptive and neutral genetic divergence, (ii) population genetic studies employing an AMOVA framework, and (iii) one illustrative example of IBA in a mosaic hybrid zone and in a phylogeographic study. For a more complete version of this table, including study design, molecular markers, and much more detailed results, see the Supporting Information

| Organism | IBA? | IBD? | Reference |
|--|--------------------------------------|------------------------------|--|
| Population genetic studies generally using distance matrices | | | |
| 1. <i>Isoodon obesulus</i> (brown bandicoot) | Yes | No | Cooper 2000 |
| 2. <i>Anolis roquet</i> (anole lizard) | Yes | No | Ogden & Thorpe 2002 |
| 3. <i>Poecilia reticulata</i> (guppies) | No | Yes | Crispo <i>et al.</i> 2006 |
| 4. <i>Canis lupus</i> (European grey wolf) | Yes | Yes | Pilot <i>et al.</i> 2006 |
| 5. <i>Coregonus clupeaformis</i> (whitefish) | Yes | N/A | Lu & Bernatchez 1999 |
| 6. <i>Timema cristinae</i> (herbivorous stick-insect) | Yes for some AFLP loci and for mtDNA | Yes (but weak) | Nosil <i>et al.</i> 2008 |
| 7. <i>Dubautia arborea</i> and <i>D. ciliolata</i> (silversword plant) | Yes | No | Friar <i>et al.</i> 2007 |
| 8. <i>Littorina saxatilis</i> (intertidal snail) | Yes | No | Grahame <i>et al.</i> 2006 |
| 9. <i>Geum urbanum</i> (forest herb) | No | No | Vandepitte <i>et al.</i> 2007 |
| 10. <i>Canis lupus</i> (North American grey wolf) | Yes | Yes | Musiani <i>et al.</i> 2007 |
| 11. <i>Hordeum spontaneum</i> (wild barley) | Yes | Yes | Owuor <i>et al.</i> 1999 |
| 12. <i>Canis lupus</i> (North American grey wolf) | Yes (for some variables) | No (but not directly tested) | Carmichael <i>et al.</i> 2007 |
| 13. <i>Alopex lagopus</i> (arctic fox) | No | No | Carmichael <i>et al.</i> 2007 |
| 14. <i>Zostera marina</i> (marine flowering plant) | No | Yes | Oetjen & Reusch 2007 |
| 15. <i>Biscutella laevigata</i> (flowering plant) | Yes | Yes | Parisod & Christin 2008 |
| 16. <i>Neochlamisus bebbinae</i> (leaf beetle) | Yes for AFLPs, No for mtDNA | No for AFLPs, Yes for mtDNA | D.J. Funk, S.P. Egan, P. Nosil, Vanderbilt University, Nashville, TN, in preparation |
| Studies employing primarily an AMOVA (or similar) framework | | | |
| 17. <i>Parus major</i> (great tit) | No | N/A | Blank <i>et al.</i> 2007 |
| 18. <i>Parus caeruleus</i> (blue tit) | No | N/A | Blank <i>et al.</i> 2007 |
| 19. <i>Osmerus mordax</i> (rainbow smelt) | No | N/A | Curry <i>et al.</i> 2004 |
| 20. <i>Loxia curvirostra</i> complex (red crossbills) | Yes | No | Parchman <i>et al.</i> 2006 |
| 21. <i>Hesperotettix viridis</i> (grasshopper) | Yes | No (but not directly tested) | Sword <i>et al.</i> 2005 |
| 22. <i>Salamandra salamandra</i> (fire salamander) | Yes | No | Steinfartz <i>et al.</i> 2007 |
| Phylogeographic and hybrid zone studies | | | |
| 23. <i>Halichoeres</i> spp. (tropical reef fish, wrasses) | Yes | No | Rocha <i>et al.</i> 2005 |
| 24. <i>Bombina</i> spp. (toads) | Yes | N/A | MacCullum <i>et al.</i> 1998; see also Vines <i>et al.</i> 2003 |

Supporting information for details). Nonetheless, the results imply that approximately 5–10% of the genome is strongly affected by divergent selection. More standardization in the reporting of future results should help to refine the estimates reported here. What this survey most clearly demonstrates is that nontrivial proportions of the genomes of disparate taxa show evidence of divergent selection. These findings are highly consistent with studies examining a few loci or particular parts of the genome (e.g. sex chromosomes), and those comparing divergence at coding vs. noncoding regions, which also regularly report heterogeneous genomic divergence (Wang *et al.* 1997;

Machado *et al.* 2002; Broughton & Harrison 2003; Hoekstra *et al.* 2004; Payseur *et al.* 2004; Llopart *et al.* 2005; Basset *et al.* 2006; Geraldles *et al.* 2006; Zayed & Whitfield 2008).

Genomic distribution of outliers To what extent are outlier loci clustered within the genome, as predicted by some theory (Gavrilets 2004, chapter 6; Kirkpatrick & Barton 2006)? The data in Table 2 suggest a range of answers, with genomic clustering of outliers ranging from low to reasonably high. Three studies located outliers on a linkage map. Two of these mapped them to many different linkage groups (13, 14 in Table 2), while the third observed a clustering of

loci associated with between-race divergence in particular regions on just a few of more than 20 chromosomes examined (6 in Table 2). Two other studies also identified the specific location of genomic differentiation. Turner *et al.* (2005) found that differentiation between forms of *Anopheles gambiae* mosquitoes involved 1.2% of the genome and was clustered into three genomic regions comprising less than 2 Mb. Clustering was somewhat weaker between house mouse (*Mus musculus*) subspecies, where differentiated regions (7.5% of the autosomal genome) were distributed among eight genomic regions from the 14–16 chromosomes examined (Harr 2006).

Four additional studies indirectly examined outlier distribution by quantifying within-population linkage disequilibrium (LD) among loci (1, 8, 9, 17 in Table 2). If outlier loci are physically linked, their LD should be elevated relative to neutral loci (Kim & Nielsen 2004), assuming the latter are widely distributed throughout the genome. Importantly, physical linkage of outliers is expected to be associated with similar levels of LD within allopatric and parapatric populations (Arnold 1992). In contrast, if LD forms between physically unlinked loci, due to migration between differentiated populations (Nei & Li 1973; Kirkpatrick *et al.* 2002), then LD will be greater where migration is higher, such as within parapatric populations (Nosil *et al.* 2006) or at the centre of a cline (Grahame *et al.* 2006). For all four relevant studies, levels of LD for outlier loci were very low and, except for one study, similar to those of nonoutlier (neutral) loci. One study did find slightly elevated LD for outlier loci relative to nonoutliers (Nosil *et al.* 2008). Because levels of within-population LD among outliers were independent from levels of between-population gene flow (migration), this study provides indirect evidence for weak physical linkage among outlier loci.

In summary, the reviewed studies provided evidence for both genomic dispersion and genomic clustering of outliers. The former observation suggests that islands of genomic divergence may be greater in number, and perhaps smaller in size, than currently thought (see also Nakazato *et al.* 2007; Mäkinen *et al.* 2008b; Turner *et al.* 2008; Wood *et al.* 2008) and is contrary to both the theoretical predictions outlined above and some empirical observations, such as quantitative genetic evidence on the tendency of genes involved in host adaptation to be sex-linked in herbivorous insects (Prowell 1998). More studies are required before explanations for these variable empirical outcomes can be evaluated, although the extent of genomic coverage (e.g. marker density) in a genome scan could contribute to this variability. We suggest that future studies report patterns of LD across different classes of loci and geographic contexts.

Comparative insights: 'ecological' genome scans The evaluation of replicated comparisons of particular types of population

pairs represents a powerful application of the genome-scan approach. This is because replicated divergence across multiple population pairs of a given type is unlikely to arise via nonselective factors such as type I error, genetic drift, or mutation rate variation (Luikart *et al.* 2003; Campbell & Bernatchez 2004; Bonin *et al.* 2006, 2007). Such studies might further allow the specific ecological causes of outlier behaviour to be identified, for example, by contrasting genetic differentiation for population pairs that are ecologically similar vs. those that are ecologically (and presumably adaptively) divergent in a specific ecological variable (Schluter & Nagel 1995; Funk 1998; Funk *et al.* 2002; Nosil 2007). The few studies making explicit comparisons between population pairs with different vs. similar ecologies (studies 6–12 in Table 2) suggest that a relatively large proportion (25–100%) of outliers are associated with divergence in a specific ecological variable. A related approach examines correlations between outliers and environmental factors (e.g. studies 5 and 18 in Table 2).

Consider the specific example of adaptation to different plant species by herbivorous insect populations. Recent studies have identified loci that are outliers in comparisons of multiple 'different-host' population pairs, but that are never outliers in comparisons of different populations that use the same host plant (Egan *et al.* 2008; Nosil *et al.* 2008). Such patterns suggest that these loci have likely evolved under the influence of host-plant-related sources of selection. By contrast, loci that are outliers only in 'same-host' population pairs are best interpreted as being affected by host-independent sources of selection (e.g. climate). For example, in a study of nine different-host and six same-host population comparisons, Egan *et al.* (2008) identified 23 outlier loci (representing 5% of all loci examined) associated exclusively with different-host population pairs of *Neochlamisus bebbianae* leaf beetles, and only five outliers associated exclusively with same-host population pairs. From this, they concluded that host-related selection plays a major role in the adaptive genomic differentiation of these populations. This same study identified three loci that were especially highly differentiated outliers in all nine different-host population comparisons and none of the same-host comparisons, thus illustrating how comparative genome scans can identify robust candidate loci (genomic regions) for further molecular characterization and evolutionary study (e.g. Wood *et al.* 2008).

Replicated divergence: adaptation via the same or different mutations? A major question in evolutionary genetics is the extent to which divergent adaptation in different geographic localities or taxa involves the same vs. different genes, alleles, and substitutions (Orr 2005). The frequency with which outliers are replicated across different population pairs in a given study provides some insight into this question. In the relevant studies, a large proportion of outlier

loci (approximately 50%) were replicated in this fashion, exhibiting outlier status across multiple population pairs (Table 2). This tendency implies that divergent adaptation may often proceed in two ways. First, it may occur by the repeated and parallel fixation of the same allele in different populations (Table 2; see also Storz & Nachman 2003; Panova *et al.* 2006; Turner *et al.* 2008). Second, an adaptive allele may arise and be locally fixed only once, followed by the subsequent spread of the ecological type carrying the new allele to multiple geographic localities.

Replicated outlier behaviour is of particular interest when divergent adaptation in different localities can be demonstrated to have evolved independently (e.g. multiple origins of ecological types), thus providing evidence for truly 'parallel' divergence (e.g. as in *Coregonus* whitefish, Campbell & Bernatchez 2004, and *Timema* walking-sticks, Nosil *et al.* 2002, 2008; Table 2). A definitive example of this process comes from threespine sticklebacks (*Gasterosteus aculeatus*), in which independent instances of adaptation to freshwater environments have repeatedly involved the loss of lateral plates (an antipredator trait) via the fixation of the same allele at the ectodysplasin gene (Colosimo *et al.* 2005). Alternatively, divergent adaptation may proceed via different mutations/loci in different localities such that particular outliers are not highly consistently observed across population comparisons (e.g. Acheré *et al.* 2005; Panova *et al.* 2006). These different mutations could either be recently derived or reflect the differential sorting of standing genetic variation (Barrett & Schluter 2008).

Genealogical discordance: population trees from neutral vs. selected loci When genetic exchange between populations varies among loci, different loci can yield different gene trees (i.e. genealogical discordance) (Maddison 1997; Via 2001; Shaw 2002; Funk & Omland 2003; Hey 2006). Loci involved in divergent adaptation and reproductive isolation are expected to reflect boundaries between biological species or ecotypes more strongly than neutral loci, because the former flow less readily between populations (Wu 2001; Dopman *et al.* 2005; Hey 2006; Xie *et al.* 2007). This process generates the prediction that selected loci are likely to group populations by adaptively relevant ecological variables, whereas neutral loci are likely to phylogenetically group populations by geographic proximity (reflecting spatial patterns of gene flow). A number of studies that have evaluated genetic structure at each of a few loci report such patterns (Beltrán *et al.* 2002; Hoekstra *et al.* 2004; Dopman *et al.* 2005; Llopart *et al.* 2005; Bull *et al.* 2006; Cano *et al.* 2006; Geraldès *et al.* 2006; Putnam *et al.* 2007; Roe & Sperling 2007). Five genome scan studies (1, 2, 4, 8, 9 in Table 2) have analogously compared population trees for data sets that either include or exclude outliers. All these studies observed that outlier loci group populations more as a function of ecology, whereas putatively neutral (nonoutlier) loci group

populations in a manner more consistent with geography (see also Via & West 2008). The extent to which outlier loci group populations according to ecology should increase as the frequency of outlier loci and population pairs exhibiting replicated divergence increases, and these studies are consistent with this prediction. Another factor is how tightly linked the outliers are to the actual targets of selection.

Genome scans incorporating QTL We conclude this section by considering studies that complement genome scans with QTL approaches. A major goal of such studies is to determine the phenotypic traits and associated genomic regions that contribute to adaptive population divergence. This can be achieved by determining which (if any) QTL from mapping experiments are also outlier loci in genome scans. Such work is relevant to heterogeneous genomic divergence because it compares levels of genetic divergence between QTL and other markers. Finding QTL that are also outlier loci puts the study of heterogeneous genomic divergence on steadier 'ecological footing', as it allows stronger inferences about underlying ecological traits and associated divergent selection than genome scans of anonymous markers alone (Stinchcombe & Hoekstra 2008). For treatment of allele frequency expectations at neutral loci vs. QTL, we refer readers to past reviews (Latta 1998, 2003; Le Corre & Kremer 2003).

To date, few studies have combined QTL and genome scan approaches. Thus, we focus on a few key examples. The first involves whitefish (*Coregonus clupeaformis*) ecotypes studied by Rogers & Bernatchez (2005, 2007). Genetic mapping identified nine QTL for phenotypic traits related to adaptation of the two ecotypes to different lake environments, while a genome scan of four sympatric pairs of ecotypes identified 24 outlier loci (among 440 loci examined). These outliers were associated with QTL, as opposed to other genomic regions, more often than expected by chance. These findings increased confidence that regions under selection had indeed been identified, and that outlier differentiation reflected divergent adaptation to lake environments (see also Ólafsdóttir *et al.* 2006). Although the authors point out that demographic, spatial, and local selective effects can influence QTL–outlier correlations (Beaumont & Balding 2004; Hahn 2006), this work clearly illustrates the inferential advantages of a more integrated approach.

The second example treats *Helianthus annuus* and *H. petiolaris* sunflowers (Yatabe *et al.* 2007). These species hybridize extensively, F₁ hybrids exhibit extremely low pollen fertility, and these fertility barriers map to chromosomal rearrangements (Rieseberg *et al.* 1999). This study, however, detected no association of outliers with QTL for morphological differences or hybrid sterility, and only weak associations of outliers with chromosomal inversions. The authors conclude that regions of differentiation between these species are very small.

A third study considers alfalfa and clover host races of *Acyrthosiphon pisum* pea aphids (Via & West 2008), and reports that outlier loci between the races are significantly clustered around QTL for traits that cause ecologically based reproductive isolation, while also demonstrating that genetic divergence decreases with increasing map distance from QTL (Fig. 3B). A final example (Mäkinen *et al.* 2008a, b) examined 103 microsatellites (many linked to known QTL) in four freshwater and three marine populations of threespine stickleback fishes. This study reported 2.8% of loci to be outliers, with the clearest signature of selection exhibited by a marker associated with the ectodysplasin gene (which codes for the ecologically important trait of lateral plate number). However, other markers associated with QTL showed no signature of selection, and two strong outliers were not associated with known QTL. These findings highlight certain limitations of this otherwise informative approach. Because QTL studies examine only a subset of phenotypic traits potentially evolving under divergent selection, outlier loci truly subject to such selection may nonetheless map to non-QTL regions (Martin *et al.* 2005). Moreover, QTL regions are often very large (tens of centimorgans) so that in the absence of extensive LD along the chromosome, genome scans might not identify loci within these regions as outliers (Cano *et al.* 2006).

Literature review 2: isolation by adaptation

Identifying relevant IBA studies We identified 22 studies relevant to the evaluation of IBA (controlling for geographic distance) from a WOS search on 'ecology and genetic and divergence', and from the studies cited by the recovered papers. In these studies, adaptive divergence was inferred using either the degree of divergence in habitat ($n = 15$) or phenotype ($n = 7$). We excluded phylogeographic studies that examine genetic divergence in relation to habitat, because they generally do not analyze genetic divergence among population pairs in relation to adaptive divergence. Similarly, studies of mosaic hybrid zones relate genetic differentiation to habitat type, but generally do not explicitly evaluate genetic divergence as a function of ecology vs. geographic distance, and have been reviewed elsewhere (Nosil *et al.* 2005). One study of each of these latter two types was nonetheless included for illustrative purposes (but excluded when calculating summary statistics). Due to the broad range of studies that could conceivably be suitable for examining IBA, we acknowledge that additional relevant studies surely exist. This contrasts with the genome scan review above, where only a more specific type of study was relevant, such that the majority of relevant studies were likely identified. Nonetheless, we have hopefully identified a representative sample of studies, and investigations using distance

matrices are likely well covered. We classified the focal studies into those that examined matrices of pairwise differences among population pairs, for example, using Mantel tests (Manly 1997), vs. those that quantified the proportion of genetic variation explained by habitat within an AMOVA framework (Excoffier *et al.* 1992). The studies treated here are assumed to evaluate loci that are neutral and not tightly linked to those under selection. Such loci represent perhaps the most intriguing and widely evaluated aspect of IBA.

Prevalence of IBA Our standard for the presence of IBA was either a significant positive correlation between adaptive and neutral genetic divergence across population pairs or significant genetic structuring between ecological types or habitats by AMOVA. With this in mind, we found evidence for IBA in 15 out of 22 studies across a variety of taxa (Table 3). Of the 16 tests employing distance matrices, 75% support IBA, as do half of the AMOVA studies. As our survey is not a formal meta-analysis, the results should not be over-interpreted. The clearest result is that IBA is not uncommon in nature.

An unresolved question is what proportion of the genome exhibits IBA. Most studies testing for IBA pool across loci to obtain population-level estimates of genetic differentiation, precluding an estimate of the proportion of the genome exhibiting IBA. However, four studies did report locus-specific results. In whitefish ecotypes (*Coregonus clupeaformis*), five of six microsatellite loci exhibited fairly strong IBA, showing correlations between ecomorphological and neutral genetic divergence that ranged from $r = 0.72$ – 0.84 (Lu & Bernatchez 1999). In host plant ecotypes of *Timema cristinae* walking-stick insects, 10% of nonoutlier AFLP loci, as well as mitochondrial DNA, exhibited significant IBA (Nosil *et al.* 2008). In host forms of *Neochlamisus bebbianae* leaf beetles, 11% of nonoutlier AFLP loci exhibited IBA, but mtDNA did not (D.J. Funk, S.P. Egan, P. Nosil, Vanderbilt University, Nashville, TN, in preparation). Finally, in wild barley (*Hordeum spontaneum*), 44% of random amplified polymorphic DNA (RAPD) loci exhibited a strong correlation with soil type (Owuor *et al.* 1999). Thus, although a genome-wide signature of IBA is not rare, the proportion of the genome exhibiting a particularly strong pattern of IBA may vary considerably. The combined trends suggest a potentially important role for natural selection in neutral genomic differentiation between populations.

The role of ecological vs. geographic factors in reducing gene flow What are the relative roles of ecology vs. geography in facilitating neutral genetic divergence? IBD was detected in 50% of studies where it was evaluated ($n = 16$), somewhat less than for IBA. Perhaps more interesting is that the presence vs. absence of IBA did not appear strongly associated with the presence vs. absence of IBD.

For example, of the 16 studies in Table 3 that tested for both IBA and IBD, six detected both, two detected neither, six detected only IBA, and two detected only IBD ($P > 0.25$, Fisher's exact test).

Consider some examples of these different types of results. Ogden & Thorpe (2002) report a primary role for ecological divergence, rather than geographic distance, in the neutral genetic divergence of a Caribbean lizard (*Anolis roquet*). Using seven microsatellite loci, genetic distance was compared among pairs of adjacent localities from three different transects, one of which cut through an ecological gradient (the 'habitat transect'), and two of which did not. Strong genetic differentiation was observed only in the habitat transect, particularly at habitat boundaries, and genetic structuring by habitat was further supported by AMOVA. The results are thus consistent with IBA. In contrast to these findings, a study of guppies (*Poecilia reticulata*) reported no evidence for IBA and a primary role for geographic distance and physical barriers to dispersal in reducing gene flow (Crispo *et al.* 2006). Finally, a study of grey wolves (*Canis lupus*) related genetic distance to both habitat type (tundra, taiga, or boreal coniferous forest) and geographic distance, reporting both IBA and IBD (Musiani *et al.* 2007). Below, we propose some hypotheses for this variability.

Hypotheses for variability among studies IBA appears relatively common, yet was not uniformly observed. What factors explain variability in IBA? An obvious possibility is that selection is sometimes too weak to cause IBA, whether via loose linkage or general barriers to gene flow (Figs 1 and 2). Another possibility is that the particular phenotypic/ecological traits evaluated are not good proxies for the major sources of divergent selection acting on study populations. Reciprocal transplant experiments indicate that divergent selection between alternative environments is very common (see Schluter 2000; Nosil *et al.* 2005 for reviews), suggesting that habitat-based indices of adaptive divergence should provide informative proxies for selection. Notably, however, in the limited sample of studies available to date, the detection of IBA appears independent from whether adaptive divergence was inferred using habitat or phenotypic data ($P > 0.25$, Fisher's exact test).

Finally, levels of gene flow can affect IBA. For example, high gene flow can overwhelm adaptive divergence (Saint-Laurent *et al.* 2003; Hendry & Taylor 2004; Smith *et al.* 2005; Crispo *et al.* 2006; Yatabe *et al.* 2007), precluding the generation of IBA. This raises the issue of reversed causality, where even if IBA is detected, levels of gene flow may be affecting the degree of adaptive divergence, rather than vice versa (Hendry *et al.* 2001; Hendry & Taylor 2004; Nosil & Crespi 2004). In addition to levels of biological gene flow, the spatial scale of sampling is also important.

At a spatial scale greater than that at which gene flow occurs, neutral divergence can occur without general barriers to gene exchange, potentially precluding the detection of IBA at this scale (even if IBA would be detectable by sampling at a smaller spatial scale).

Studies combining tests for outlier loci with the examination of IBA

As described above, divergent selection can cause heterogeneous genomic divergence via different processes. Questions remain concerning how these different effects of selection interact, because outlier loci and IBA have rarely been evaluated within the same system. Here, we summarize findings from the few existing examples.

One of the first applications of a genome scan to natural populations of a nonmodel organism was that of Wilding *et al.* (2001), who identified AFLP loci that were outliers between upper and lower shore ecotypes of *Littorina saxatilis* snails. A subsequent study further examined these outliers, as well as the effects of adaptive divergence on neutral differentiation (Grahame *et al.* 2006). These studies showed outlier loci to form sharp clines in allele frequencies in transects between the upper and lower shore (Fig. 5A). LD among outlier loci was low except in the middle of the cline, as expected when migration between differentiated populations generates LD between unlinked loci (see also Wood *et al.* 2008). Nonoutlier loci exhibited higher F_{ST} between ecotypes relative to within-ecotype comparisons (independent of geographic distance; Fig. 5B). These results are consistent with selection acting on unlinked outlier loci and creating a general barrier to gene flow across a contact zone, resulting in IBA at neutral loci.

Another combined analysis was conducted by Ólafsdóttir *et al.* (2006), who examined both QTL putatively under selection and neutral markers. These authors reported elevated divergence between stickleback (*Gasterosteus aculeatus*) ecotypes at microsatellites associated with QTL relative to putatively neutral non-QTL markers. They also observed IBA at neutral loci as a function of divergent predation regimes.

Finally, a study on a herbivorous insect used AFLPs to identify outliers and evaluate IBA among populations associated with either of two host plants (Nosil *et al.* 2008). In an investigation of *Timema cristinae* walking-stick insects 8% of AFLP loci proved to be outliers and 1–2% of loci were identified as being subject to host-plant-related selection, using the comparative approach described above. This study found little evidence for IBA when loci were pooled (Fig. 5C). However, as described earlier, locus-specific analyses revealed that 10% of non-outlier loci exhibited IBA, as did mitochondrial DNA (Fig. 5D). In summary, patterns for outlier loci and IBA exhibit both similarities and differences among systems.

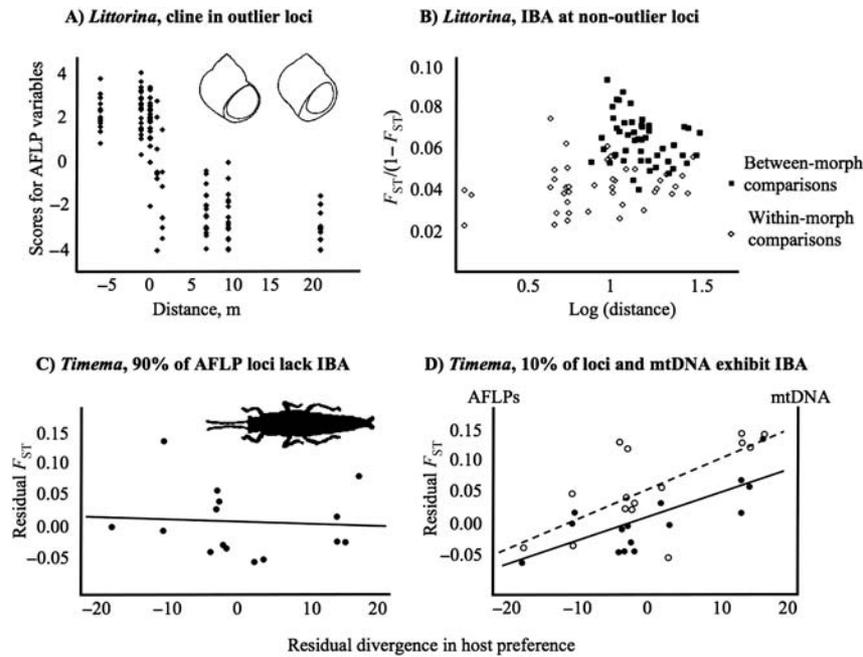


Fig. 5 Patterns of outlier differentiation and IBA in two study systems where both have been examined. These examples illustrate the heterogeneous nature of genomic differentiation: outlier loci exhibit greater divergence than nonoutlier loci (A, B), and differentiation varies among nonoutlier loci such that (only) a fraction of them exhibit strong IBA (C–F). (A) A cline in allele frequencies for outlier loci from comparisons of upper and lower shore ecotypes of *Littorina saxatilis* snails. (B) Greater neutral (i.e. non-outlier) differentiation between than within the *L. saxatilis* ecotypes, for a given geographic distance, illustrating IBA. Modified from Grahame *et al.* (2006). (C) Lack of association between adaptive phenotypic divergence and genetic differentiation at 90% of AFLP loci between host plant ecotypes of *Timema cristinae*. (D) Evidence for IBA in *T. cristinae* at 10% of AFLP loci (filled circles, solid line) and mtDNA (open circles, dashed line) based on analyses controlling for geographic distance. Modified from Nosil *et al.* (2008). Both the *Littorina* and *Timema* figures are reprinted with permission of the Society for the Study of Evolution.

The growth of genomic islands of divergence: alternative models

Our literature reviews detected some general trends, but also variability in patterns of outlier behaviour and of IBA. This variability may reflect factors determining the size and number of differentiated regions (i.e. genomic islands of divergence) across the genome. As described, the effects of divergent selection can extend across a chromosome via different effects on tightly linked, loosely linked, and unlinked loci. The further the effects of divergent selection extend along chromosomes, the larger the associated regions of elevated differentiation. The stronger the selective barriers to gene flow, the greater the number of loci expected to exhibit elevated differentiation via processes that contribute to IBA. But how do genomic islands of divergence (hereafter 'islands' for simplicity) grow? Here, we present models for island growth (see also Fig. 4, Table 1, and the Supporting information).

Model I. Allopatric model

Genetic divergence and reproductive isolation during allopatric differentiation is unimpeded by gene flow and

increases with time (reviewed in Coyne & Orr 2004). Thus, the allopatric model predicts that the number of islands should be positively correlated with time since population divergence. As compared to models of divergence with gene flow (see below), regions of differentiation are not predicted to be as highly clustered within the genome during allopatric divergence, because divergence at all regions (i.e. not only those already exhibiting reduced introgression) can proceed unimpeded by gene flow. An empirical study comparing an allopatric species pair of *Drosophila* to a sympatric pair supports this prediction (Brown *et al.* 2004). This raises the issue of the degree to which low clustering of differentiated regions within the genome represents a genetic signature of allopatric divergence. Another pattern that may be associated with this model is high baseline levels of neutral differentiation, because physical barriers to gene flow make divergence via genetic drift likely (i.e. even without the evolution of 'general barriers' to gene flow). Also, islands might be relatively small owing to the absence of certain conditions characterizing island growth in the other models. Nonetheless, even in the allopatric case, regions under stronger selection are expected to produce islands of greater size and height through their stronger effects on selected

and linked neutral regions within each population, and islands may reach nontrivial height since adaptive divergence is unconstrained by gene flow. Thus, the allopatric model predicts many small to modest-sized islands, whose number and height are a positive function of time and selection strength (Yatabe *et al.* 2007), and which are distributed throughout the genome, as recently observed between allopatric populations of ferns (Nakazato *et al.* 2007).

Model II. Ecological (divergent selection) model with gene flow

The growth of islands in the face of gene flow is likely to differ from their growth in allopatric conditions, with the effects of time being less clear. Genes under divergent selection, and those tightly linked to them, will experience reduced introgression relative to neutral, unlinked loci. The chance that a new mutation (whether adaptive or neutral) will persist and increase in frequency is highest in regions of reduced introgression (Gavrilets 2004, chapter 6). Therefore, differentiated loci are expected to accumulate in genomic regions that already harbour genes under divergent selection, leading to increases in the number of genes within an island, and thus, in island size, and potentially generating a positive feedback loop. This model thus predicts that genes affecting local adaptation will form clusters within the genome rather than being more evenly distributed across it. QTL studies demonstrating that different adaptive traits map to similar genomic regions, but not within known inversions, are consistent with this prediction (e.g. *Acyrtosiphon* pea aphids, Hawthorne & Via 2001; *Heliconius* mimetic butterflies, Kronforst *et al.* 2006; *Coregonus* whitefish ecotypes, Rogers & Bernatchez 2007; *Gasterosteus* sticklebacks, Albert *et al.* 2007), although pleiotropy could also contribute to these results.

A number of issues will affect the generality of the process outlined above. A major one is whether divergent selection typically persists long enough for new mutations to arise and be captured by regions of reduced introgression. Empirical evidence suggests that anciently diverged, sympatric species that have presumably been subject to divergent selection for extended periods of time sometimes exhibit very small regions of differentiation (Mallet *et al.* 2007; Yatabe *et al.* 2007), but in other times exhibit larger ones (Llopart *et al.* 2005). Why do islands sometimes appear to grow, and other times do not? Some of this variability may be due to how many different islands (i.e. gene regions) affect a given trait under divergent selection, with more islands perhaps associated with more opportunity to capture new mutations within a least one island, thereby resulting in island growth. A counterargument is that more genes (i.e. islands) affecting a trait can result in

weaker per-locus selection coefficients (Gavrilets & Vose 2005), thereby constraining divergence with gene flow.

Another issue is the capacity of natural selection to favour the evolution of tighter linkage among loci, for example, to keep beneficial genotypic combinations together (Kimura 1956; Bodmer & Parsons 1962; Kojima & Schaffer 1964; Butlin 2005). This process could proceed via the evolution of modifier loci that suppress recombination (Kouyos *et al.* 2006) and facilitate the growth of genomic islands. Support for the evolution of tighter linkage is provided by the evolution of 'supergenes', that is, groups of neighbouring genes on a chromosome that are inherited together. Different genes within a supergene tend to affect different, if sometimes related, traits (Nabours *et al.* 1933; Sheppard 1953; Clarke & Sheppard 1960; Turner 1967a; Sinervo & Svensson 2002). Classic examples are genes affecting different colour-pattern traits in insects, such as instances where the fitness of an allele at a locus affecting one colour-pattern element depends on which alleles are present at a different locus affecting another colour-pattern element (Nabours *et al.* 1933; Sheppard 1953). Such a scenario generates selection favouring the retention of particular genotypic combinations (i.e. tighter linkage; Nabours *et al.* 1933; Sheppard 1953; Clarke & Sheppard 1960; Turner 1967a; Sinervo & Svensson 2002), but can increase linkage only within chromosomes (Turner 1967b; Charlesworth & Charlesworth 1975; Joron *et al.* 2006). In sum, compared to the allopatric model, the ecological model predicts greater genomic clustering of genomic regions under selection and fewer islands. The ecological model further predicts that islands can be small, but will sometimes be large and include multiple and sometimes interacting selected genes.

Model III. Structural model with gene flow

The origin and growth of sizeable islands might be further facilitated by the structural organization of the genome, notably by chromosomal inversions (Noor *et al.* 2001, 2007; Rieseberg 2001; Ortiz-Barrientos *et al.* 2002; Butlin 2005; Machado *et al.* 2007). Inversions themselves might cause postmating isolation, but could further promote differentiation by facilitating adaptive divergence via various processes. The 'protection from introgression' hypothesis proposes that inversions are initially established by some unknown mechanism, perhaps in allopatry. When the inversion-bearing populations come into contact, the inversions promote adaptive divergence (and associated genomic divergence) by reducing introgression at large regions of the genome and protecting favourable genotypic combinations that arise within these regions from being broken up by recombination (Rieseberg 2001; Noor *et al.* 2001; Brown *et al.* 2004). The related 'inversions first' scenario posits that once inversions are established, genetic

differences between taxa can easily build up within them (Navarro & Barton 2003). Finally, the recent 'selective spread' hypothesis posits that a newly formed inversion captures locally adapted alleles at two or more loci in hybridizing populations (Kirkpatrick & Barton 2006; Manoukis *et al.* 2008). These co-occurring loci confer a fitness advantage to the inversion by keeping well-adapted genotypes intact, thereby allowing the inversion to spread via selection. This scenario differs from the others in positing that locally adapted alleles within an inversion are the cause of the inversion's spread, rather than a consequence of its existence.

The extent to which inversions promote island growth under these scenarios depends on two main factors. The first is how frequently genes under selection or promoting reproductive isolation occur within inversions. The second factor is the extent to which inversions reduce introgression at regions outside of them. Recent work reported that levels of genetic differentiation between *Drosophila* species were indeed somewhat elevated just outside of inversions, but dropped off markedly even just a few megabases outside the inversion (Machado *et al.* 2007; Noor *et al.* 2007).

Despite its simplicity and potentially great explanatory capacity, empirical evidence on the importance of the structural model appears mixed. On one hand, examples exist where inversions appear to have promoted genomic divergence and speciation (Rieseberg *et al.* 1999; Noor *et al.* 2001, 2007; Rieseberg 2001; Feder *et al.* 2003a, b; Butlin 2005; Basset *et al.* 2006; Machado *et al.* 2007; Yatabe *et al.* 2007; Manoukis *et al.* 2008). On the other hand, our genome scan review found that islands were often genomically scattered rather than clustered, contrary to expectation if such regions tend to reside within inversions (Table 1). Moreover, of the two studies explicitly focusing on islands of divergence (Turner *et al.* 2005; Harr 2006), only the first found regions of high differentiation to be associated with regions of reduced recombination (specifically, centromeres, Fig. 4B). Thus, it appears that although inversions can promote island formation and growth, they are not required.

The structural and ecological models are similar in some aspects, but differ most clearly in two respects. First, a smaller number and greater size of genomic islands is predicted for the structural model because inversions are often not plentiful within genomes, but commonly consume extensive regions of chromosomes. Second, the structural model may allow longer evolutionary persistence of islands, due to the strong inhibition of recombination caused by inversions.

Integration of the different models

The models above are not mutually exclusive, and might interact. For example, the ecological and structural models

could operate simultaneously, resulting in clustering of genes affecting local adaptation within (and around) chromosomal inversions, as predicted by recent theory (Kirkpatrick & Barton 2006). Another possible interaction is between the allopatric model and models involving gene flow, given that the often extended process of speciation can involve multiple geographic modes. For example, some divergence might occur in allopatry and some in sympatry (Feder *et al.* 2003a; Rundle & Schluter 2004; Rundle & Nosil 2005; Xie *et al.* 2007). Such a geographically pluralistic view of speciation has implications for the heterogeneity of genomic divergence observed. For example, the amount of adaptive divergence and reproductive isolation that evolves during an initial allopatric period will influence patterns of genomic differentiation following secondary contact. If very little reproductive isolation evolved, then widespread gene flow will ultimately erode accumulated differentiation at islands not possessing the introgression-resisting characteristics described above. If strong reproductive isolation evolved, then accumulated differentiation might largely be maintained, and further divergence can occur. Thus, the number of genomic islands should be positively associated with the degree of divergence during the allopatric period, itself a function, in part, of the duration of this period. A related point is that the age of a contact (e.g. hybrid) zone will affect patterns of heterogeneous genomic divergence, because it takes time for gene flow to erode divergence upon secondary contact (Strasburg & Rieseberg 2008). Thus, the size and number of islands might decrease through time since secondary contact.

Factors affecting all models

A number of additional factors are relevant to island growth under all models. One is the genomic distribution of genes subject to divergent selection, as island formation and growth will be facilitated by any tendency for selected genes to occur in physical proximity. Another is the distribution of linkage disequilibrium (LD) and its rate of decay with increasing distance from selected regions (Charlesworth *et al.* 1997; Nielsen 2005; Cano *et al.* 2006). The more rapidly LD decays, the smaller the resulting islands. The relevance of this issue is indicated by evidence that levels of LD can vary among genomic regions and taxa, for example, due to variation in recombination rates (Reich *et al.* 2001; Scotti-Saintagne *et al.* 2004; Liu & Burke 2006; Arunyawat *et al.* 2007). In wild sunflower (*Helianthus annuus*), for instance, LD among loci falls to negligible levels within 200 bp, whereas in cultivated varieties of this species, it extends up to 1100 bp (Liu & Burke 2006). Another factor is whether forms of reproductive isolation (e.g. assortative mating) evolve, with genomic differentiation via selection being facilitated when such

barriers to gene flow evolve. Reproductive isolation itself can involve epistatic interactions between different loci, perhaps even loci on different chromosomes. The effects of such epistasis on the nature of islands are poorly understood, illustrating a need for work on 'speciation' islands.

A factor of general importance, as noted in our initial description of islands, is the strength of selection, with stronger selection leading to stronger hitchhiking effects (Fig. 1A), and thus, presumably to larger and higher islands. Thus, islands may be expected to grow through time as populations in alternative environments gradually become more divergently adapted, resulting in stronger selection against immigrants. Relatedly, patterns of island growth might vary across different stages of the speciation process (for consideration of the speciation continuum, see Wu 2001; Berlocher & Feder 2002; Drès & Mallet 2002; Hey *et al.* 2003; Coyne & Orr 2004; De Queiroz 2005; Funk *et al.* 2006; Rueffler *et al.* 2006; Mallet *et al.* 2007; Svensson *et al.* 2008; Nosil & Sandoval 2008). Many of the examples in our genome scan review (Table 2) compare conspecific populations, rather than distinct species. Thus, it is possible that our findings reflect a bias towards the earlier stages of speciation, when islands may still be small, reflecting the few regions under strong selection and low levels of associated reproductive isolation. The later stages of speciation might be characterized by different types of divergence, for example, larger islands that contain inversions that facilitate the long-term persistence of differentiation. Studies of population/species pairs spanning the range of divergences that cumulatively represent the speciation continuum might contribute to a more comprehensive understanding of how divergent selection affects heterogeneous genomic divergence.

Conclusions and future directions

Divergent selection is predicted to yield heterogeneous divergence across the genome. And indeed, recent work is beginning to document variation in the prevalence, distribution, and size of differentiated genomic regions. A variety of outstanding questions exist about the causes and consequences of these patterns. Here, we have attempted to integrate emerging ideas and findings from different research traditions in order to facilitate future exploration of the role of divergent selection in the evolution of heterogeneous genomic divergence, including the testing of associated predictions (Table 1). A major conclusion is that divergent selection plays multiple roles. These include its effects on fitness-associated loci, its effects on both tightly and loosely linked neutral loci via hitchhiking, its facilitation of genetic drift by countering gene flow, and its influence on the growth of genomic islands. These sometimes nonintuitive contributions illustrate the capacity

of selection to affect many aspects of genomic differentiation. More specifically, the modest data collected so far illustrate considerable variability in the number, size, and genomic distribution of strongly differentiated regions. Avenues for future research should include, first and foremost, the collection of more data from more taxa. This is clearly required for generalities to more fully emerge and to distinguish between alternative hypotheses. Also important will be evaluating how particular ecological, genetic, and geographic factors help explain observed variability. Finally, increased integration of different methods and the incorporation of additional methods (e.g. gene expression analysis, Derome *et al.* 2006; Roberge *et al.* 2007) will allow new questions to be addressed. We hope that the ideas and data reviewed here help promote the advancement of the emerging field of population genomics.

Acknowledgements

We thank B. Crespi, A. Hendry, J. Galindo, J. Lee-Yaw, H. Collin, S. Rogers, and D. Schluter for discussions about genomic divergence and speciation. S. Rogers, S.P. Egan, J. Galindo, L. Bernatchez, and three anonymous reviewers provided many useful comments on earlier versions of this manuscript. A. Posada provided assistance with the artwork. During the preparation of this manuscript, P.N. was funded by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council (NSERC) Canada, D.J.F. was funded by Vanderbilt University and by NSF grants IOB0447659 and DEB0640741, and D.O.B. was funded by The University of Queensland.

Glossary

Heterogeneous genomic divergence – variation among genomic regions, of any given size or content, in their degree of population differentiation

Divergent selection – selection that acts in contrasting directions between two populations, including the special case where selection favours both extremes within a single population (i.e. disruptive selection)

F_{ST} – a measure of population differentiation based on the proportion of genetic variation that occurs between populations vs. within populations

Outlier loci – loci whose level of population differentiation statistically exceeds neutral expectations, where neutral expectations are often inferred using simulations

Loose linkage – weak physical linkage along a chromosome of a neutral locus to a selected locus that is sufficient to yield somewhat elevated genetic differentiation via genetic hitchhiking, but not strong enough differentiation to yield outlier status

Isolation by adaptation (IBA) – a positive correlation between the degree of adaptive phenotypic divergence between populations and their level of molecular genetic differentiation, independent from geographic distance (IBA involving neutral loci that are not tightly linked to those under divergent selection is of special interest)

Genomic island of divergence – a region of the genome, of any size, whose divergence exceeds neutral expectations

References

- Achere V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Molecular Ecology*, **14**, 3191–3201.
- Albert AYK, Sawaya S, Vines TH *et al.* (2007) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*, **62**, 76–85.
- Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics and Development*, **11**, 635–641.
- Arnold SJ (1992) Constraints on phenotypic evolution. *American Naturalist*, **140**, S85–S107.
- Arnqvist G, Rowe L (2005) *Sexual Conflict*. Princeton University Press, Princeton, New Jersey.
- Arunyawat U, Stephan W, Stadler T (2007) Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Molecular Biology and Evolution*, **24**, 2310–2322.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.
- Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.
- Barrett R, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.
- Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341–359.
- Barton NH (1983) Multilocus clines. *Evolution*, **37**, 454–471.
- Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**, 1553–1562.
- Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity*, **57**, 357–376.
- Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.
- Basset P, Yannic G, Brunner H, Hausser J (2006) Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. *Evolution*, **60**, 1718–1730.
- Bateson W (1909) Heredity and variation in modern lights. In: *Darwin and Modern Science* (ed. Seward AC), pp. 85–101. Cambridge University Press, Cambridge, UK.
- Beaumont MA (2005) Adaptation and speciation: what can F_{ST} 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.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.
- Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population genomics: whole genome analysis of polymorphism and divergence in *Drosophila simulans*. *Public Library of Science, Biology*, **511**, e310.
- Beltrán M, Jiggins CD, Bull V *et al.* (2002) Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution*, **19**, 2176–2190.
- 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, UK.
- Berlacher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, **47**, 773–815.
- Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.
- Blank J, Stauss MJ, Tomiuk J, Fietz J, Segelbacher G (2007) Habitat type does not affect population genetic structure in sympatric great tits (*Parus major*) and blue tits (*P. caeruleus*). *Journal of Negative Results*, **4**, 1–14.
- Bodmer WF, Parsons PA (1962) Linkage and recombination in evolution. *Advances in Genetics*, **11**, 1–100.
- Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionist. *Molecular Ecology*, **16**, 3737–3758.
- Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.
- Bowcock AM, Kidd JR, Mountain JL *et al.* (1991) Drift, admixture and selection in human evolution — a study with DNA polymorphisms. *Proceedings of the National Academy of Sciences, USA*, **88**, 839–843.
- Broughton RE, Harrison RG (2003) Nuclear gene genealogies reveal historical, demographic, and selective factors associated with speciation in field crickets. *Genetics*, **163**, 1389–1401.
- Brown KM, Burk LM, Henagan LM, Noor MA (2004) A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution*, **58**, 1856–1860.
- Bull V, Beltran M, Jiggins CD, McMillan WO, Bermingham E, Mallet J (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biology*, **21**, 11.
- Bulmer MG (1972) Multiple niche polymorphism. *The American Naturalist*, **106**, 254–257.
- Butlin RK (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.
- Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.
- Cano JM, Matsuba C, Mäkinen H, Merilä J (2006) The utility of QTL-linked markers to detect selective sweeps in natural populations — a case study of the Eda gene and a linked marker in threespine stickleback. *Molecular Ecology*, **15**, 4613–4621.
- Carmichael LE, Krizan J, Nagy JA *et al.* (2007) Historical and ecological determinants of genetic structure in arctic candids. *Molecular Ecology*, **16**, 3466–3483.
- Charlesworth D, Charlesworth B (1975) Theoretical genetics of Bayesian mimicry II. *Evolution of Supergenes Journal of Theoretical Biology*, **55**, 305–324.
- Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.
- Clarke CA, Sheppard PM (1960) Supergenes and mimicry. *Heredity*, **14**, 175–185.
- Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **310**, 1928–1933.
- Cooper ML (2000) Random amplified polymorphic DNA analysis of southern brown bandicoot (*Isodon obesulus*) populations in Western Australia reveals genetic differentiation related to environmental variables. *Molecular Ecology*, **9**, 469–479.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, Massachusetts.

- Crespi BJ (2007) Sly FOXP2: genomic conflict in the evolution of language. *Trends in Ecology & Evolution*, **22**, 174–175.
- Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, **15**, 49–62.
- Curry RA, Currie SL, Bernatchez L, Saint-Laurent R (2004) The rainbow smelt, *Osmerus mordax*, complex of Lake Utopia: threatened or misunderstood? *Environmental Biology of Fishes*, **69**, 153–166.
- De Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of National Academy of Sciences, USA*, **102**, 6600–6607.
- Derome N, Duchesne P, Bernatchez L (2006) Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchell) ecotypes. *Molecular Ecology*, **15**, 1239–1249.
- Dettman JR, Sirjusingh C, Kohn LM, Anderson JB (2007) Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature*, **447**, 585–588.
- Dobzhansky T (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, **121**, 113–125.
- Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.
- Dobzhansky T (1951) *Genetics and the Origin of Species*, 3rd edn. Columbia University Press, New York.
- Dopman EB, Perez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences, USA*, **102**, 14706–14711.
- Drès M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **357**, 471–492.
- Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host-association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, **62**, 1162–1181.
- Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings Royal Society B: Biological Sciences*, **271**, 97–105.
- Endler JA (1973) Gene flow and population differentiation. *Science*, **179**, 243–250.
- Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Faubet P, Gaggiotti OE (2008) A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics*, **178**, 1491–1504.
- Feder JL, Berlocher SH, Roethele JB *et al.* (2003a) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences, USA*, **100**, 10314–10319.
- Feder JL, Roethele FB, Filchak K, Niedbalski J, Romero-Severson J (2003b) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics*, **163**, 939–953.
- Felsenstein J (1976) The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, **10**, 253–280.
- Felsenstein J (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, **35**, 124–138.
- Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, UK.
- Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.
- 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.
- Friar EA, Cruse-Sanders JM, McGlaughlin EM (2007) Gene flow in *Dubautia arborea* and *D. ciliolata*: the roles of ecology and isolation by distance in maintaining species boundaries despite ongoing hybridization. *Molecular Ecology*, **16**, 4028–4038.
- Funk DJ (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution*, **52**, 1744–1759.
- Funk DJ, Omland KE (2003) The frequency, causes and consequences of species level paraphyly and polyphyly: insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **34**, 397–423.
- Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica*, **116**, 251–267.
- Funk DJ, Nosil P, Etges B (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences, USA*, **103**, 3209–3213.
- Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, New Jersey.
- Gavrilets S, Cruzan MB (1998) Neutral gene flow across single locus clines. *Evolution*, **52**, 1277–1284.
- Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences, USA*, **102**, 18040–18045.
- Geraldes A, Ferrand N, Nachman NW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, **173**, 919–933.
- 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.
- Hahn MW (2006) Accurate inference and estimation in population genomics. *Molecular Biology and Evolution*, **23**, 911–918.
- Haig D (2004) Genomic imprinting and kinship: how good is the evidence? *Annual Review of Genetics*, **38**, 553–585.
- Haldane JBS (1930) A mathematical theory of natural and artificial selection. Part IV. Isolation. *Proceedings of the Cambridge Philosophical Society*, **26**, 220–230.
- Haldane JBS (1932) *The Causes of Evolution*. Longmans, New York.
- Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research*, **16**, 730–737.
- Harrison RG (1991) Molecular changes at speciation. *Annual Review of Ecology and Systematics*, **22**, 281–308.
- Hartl DL, Clark AG (1997) *Principles of Population Genetics*, 3rd edn. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, **412**, 904–907.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.
- Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology and Systematics*, **37**, 67–93.

- Hendry AP (2004) Selection against migrants contributes to the rapid evolution of ecologically-dependent reproductive isolation. *Evolutionary Ecology Research*, **6**, 1219–1236.
- Hendry AP, Day T (2005) Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology*, **14**, 901–906.
- Hendry AP, Taylor EB (2004) How much of the variation in adaptive divergence can be explained by gene flow? – an evaluation using lake-stream stickleback pairs. *Evolution*, **58**, 2319–2331.
- Hendry AP, Day T, Taylor EB (2001) Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution*, **55**, 459–466.
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics and Development*, **16**, 592–596.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.
- Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution*, **18**, 597–603.
- Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genetical Research*, **8**, 269–294.
- Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in neutral and selected genes. *Evolution*, **58**, 1329–1341.
- Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.
- Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach of adaptation. *Molecular Ecology*, **16**, 3955–3969.
- Joron M, Papa R, Beltrán M *et al.* (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *Public Library of Science, Biology*, **4**, e303.
- Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology*, **15**, 3469–3480.
- Kaplan NL, Hudson RR, Langley CH (1989) The 'hitchhiking effect' revisited. *Genetics*, **123**, 887–899.
- Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167**, 1513–1524.
- Kimura M (1956) A model of genetic system which leads to closer linkage by natural selection. *Evolution*, **10**, 278–287.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624–626.
- Kimura M (1986) DNA and the neutral theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **312**, 343–354.
- King JL, Jukes TH (1969) Non-Darwinian evolution. *Science*, **164**, 788–798.
- Kirkpatrick M, Barton NH (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434.
- Kirkpatrick M, Johnson T, Barton N (2002) General models of multilocus evolution. *Genetics*, **161**, 1727–1750.
- Kojima K, Schaffer HE (1964) Accumulation of epistatic gene complexes. *Evolution*, **18**, 127–129.
- Kouyos RD, Otto SP, Bonhoeffer S (2006) Effect of varying epistasis on the evolution of recombination. *Genetics*, **173**, 589–597.
- Kronforst MR, Young LG, Kapan DD, McNeely C, O'Neill RJ, Gilbert LE (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences, USA*, **103**, 6575–6580.
- Lande R (1992) Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution*, **46**, 381–389.
- Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *The American Naturalist*, **151**, 283–292.
- Latta RG (2003) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist*, **161**, 51–58.
- Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*, **164**, 1205–1219.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*, **173**, 321–330.
- Llort A, Lachaise D, Coyne JA (2005) Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics*, **171**, 197–210.
- Lu GQ, Bernatchez L (1999) Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution*, **53**, 1491–1505.
- 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, Nürnberger B, Barton NH, Szymura JM (1998) Habitat preference in *Bombina* hybrid zone in Croatia. *Evolution*, **52**, 227–239.
- Machado C, Kliman RM, Markert JM, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and its close relatives. *Molecular Biology and Evolution*, **19**, 472–488.
- Machado CA, Haselkorn TS, Noor MAF (2007) Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **175**, 1289–1306.
- Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.
- Mäkinen HS, Cano JM, Merilä J (2008a) Identifying footprints of directional and balancing selection in marine and freshwater threespine stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology*, **17**, 3565–3582.
- Mäkinen HS, Shikano T, Cano JM, Merilä J (2008b) Hitchhiking mapping reveals a candidate genomic region for natural selection in three-spined stickleback chromosome VIII. *Genetics*, **178**, 435–465.
- Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology & Evolution*, **10**, 294–299.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.
- Mallet J (2006) What has *Drosophila* genetics revealed about speciation? *Trends in Ecology & Evolution*, **21**, 186–193.
- Mallet J, Barton NH (1989) Strong natural selection in a warning-color hybrid zone. *Evolution*, **43**, 421–431.
- Mallet J, Beltran M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology*, **7**, 28.

- Manly BFJ (1997) *Randomization and Monte Carlo Methods in Biology*, 2nd edn. Chapman & Hall, New York.
- Manoukakis NC, Powell JR, Toure MB *et al.* (2008) A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *Proceedings of the Royal Society B: Biological Sciences*, **105**, 2940–2945.
- Martin NH, Bouck AC, Arnold ML (2005) Loci affecting long-term hybrid survivorship in Louisiana irises: implications for reproductive isolation and introgression. *Evolution*, **59**, 2116–2124.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetical Research Cambridge*, **23**, 23–35.
- Mayr E (1947) Ecological factors in speciation. *Evolution*, **1**, 263–288.
- Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, UK.
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, **17**, 285–291.
- Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.
- Miller NJ, Ciosi M, Sappington TW, Ratcliffe ST, Spencer JT, Guillemaud T (2007) Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. *Journal of Applied Entomology*, **131**, 378–385.
- Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.
- Muller HJ (1940) Bearings of the *Drosophila* work on systematics. In: *The New Systematics* (ed. Huxley JS), pp. 185–268. Clarendon Press, Oxford, UK.
- Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biological Symposia*, **6**, 71–125.
- 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.
- Musiani M, Leonard JA, Cluff HD *et al.* (2007) Differentiation of tundra/taiga and boreal coniferous forest wolves: genetics, coat colour and association with migratory caribou. *Molecular Ecology*, **16**, 4149–4170.
- Nabours RK, Larson I, Hartwig N (1933) Inheritance of color patterns in the grouse locust *Acrydium arenosum* Burmeister. (Tettigidae). *Genetics*, **18**, 159–171.
- Nachman MW (2006) Detecting selection at the molecular level. In: *Evolutionary Genetics, Concepts and Case Studies* (eds Fox CW, Wolf JB), pp. 103–118. Oxford University Press, Oxford, UK.
- Nakazato T, Jung M-K, Housworth EA, Rieseberg KH, Gastony GJ (2007) A genome wide study of reproductive barriers between allopatric populations of a homosporous fern, *Ceratopteris richardii*. *Genetics*, **177**, 1141–1150.
- Navarro A, Barton NH (2003) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, **57**, 447–459.
- Nei N, Li W-H (1973) Linkage disequilibrium in subdivided populations. *Genetics*, **75**, 213–219.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. *Nature Review of Genetics*, **7**, 851–861.
- Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences, USA*, **98**, 12084–12088.
- Noor MAF, Garfield DA, Schaeffer SW, Machado CA (2007) Divergent between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics*, **177**, 1417–1428.
- Nosil P (2007) Divergent host-plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking-sticks. *American Naturalist*, **169**, 151–162.
- Nosil P (2008) Speciation with gene flow may be common. *Molecular Ecology*, **17**, 2103–2106.
- Nosil P, Crespi BJ (2004) Does gene flow constrain trait divergence or vice-versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution*, **58**, 101–112.
- Nosil P, Sandoval CP (2008) Ecological niche dimensionality and the evolutionary diversification of stick insects. *PLoS ONE*, **3**, e1907.
- Nosil P, Crespi BJ, Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, **417**, 440–443.
- Nosil P, Vines TH, Funk DJ (2005) Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.
- Nosil P, Crespi BJ, Sandoval CP, Kirkpatrick M (2006) Migration and the genetic covariance between habitat preference and performance. *The American Naturalist*, **167**, E66–E78.
- Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: ‘isolation-by-adaptation’ and multiple roles for divergent selection. *Evolution*, **62**, 316–336.
- O’Malley KG, Camara MD, Banks MA (2007) Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, **16**, 4930–4941.
- 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**, 5156–5167.
- Ogden R, Thorpe RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences, USA*, **99**, 13612–13615.
- Ohta T (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, **23**, 263–286.
- Ohta T (2002) Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences, USA*, **99**, 16134–16137.
- Ólafsdóttir GÁ, Snorrason SS, Ritchie MG (2006) Morphological and genetic divergence of intralacustrine stickleback morphs in Iceland: a case for selective differentiation? *Journal of Evolutionary Biology*, **20**, 603–616.
- Orr HA (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, **139**, 1805–1813.
- Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.
- Orr HA, Turelli M (2001) The evolution of postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities. *Evolution*, **55**, 1085–1094.
- Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. *Current Opinion in Genetics and Development*, **14**, 675–679.

- Ortiz-Barrientos D, Reiland J, Hey J, Noor MAF (2002) Recombination and the divergence of hybridizing species. *Genetica*, **116**, 167–178.
- Owuor ED, Fahima T, Beharav A, Korol A, Nevo E (1999) RAPD divergence caused by microsite edaphic selection in wild barley. *Genetica*, **105**, 177–192.
- Panova M, Hollander J, Johannesson K (2006) Site-specific genetic divergence in parallel hybrid zones suggests nonallopatric evolution of reproductive barriers. *Molecular Ecology*, **15**, 4021–4031.
- Parchman TL, Benkman CW, Britch SC (2006) Patterns of genetic variation in the adaptive radiation of New World crossbills (Aves: Loxia). *Molecular Ecology*, **15**, 1873–1887.
- Parisod C, Christin P-A (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist*, **178**, 436–447.
- Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.
- 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.
- Pilot M, Jedrzejewski W, Branicki W *et al.* (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology*, **15**, 4533–4553.
- Presgraves DC (2007) Does genetic conflict drive molecular evolution of nuclear transport genes in *Drosophila*? *Bioessays*, **29**, 386–391.
- Presgraves DC, Balagopalan L, Abmayr SA, Orr HA (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature*, **243**, 715–719.
- Prowell DP (1998) Sex linkage and speciation in Lepidoptera. In: *Endless Forms: Species and Speciation* (eds Berlocher S, Howard D), pp. 309–319. Oxford Press, New York.
- Putnam AS, Scriber JM, Andolfatto P (2007) Discordant divergence times among Z chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution*, **61**, 912–927.
- Reich DE, Cargill M, Bolk S *et al.* (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.
- Rice WR (1998) Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In: *Endless Forms: Species and Speciation* (eds DJ Howard, SH Berlocher), pp. 261–270. Oxford University Press, New York.
- Riebler A, Held L, Stephan W (2008) Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics*, **178**, 1817–1829.
- Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.
- Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two wild sunflower species. *Genetics*, **152**, 713–727.
- Roberge C, Guderley H, Bernatchez L (2007) Genome-wide identification of genes under selection: gene transcription Qst Scan in diverging Atlantic salmon subpopulations. *Genetics*, **177**, 1011–1022.
- Robertson A (1961) Inbreeding in artificial selection programmes. *Genetical Research*, **2**, 189–194.
- Rocha LA, Robertson DR, Roman J, Bowen BW (2005) Ecological speciation in tropical reef fishes. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 573–579.
- Roe AD, Sperling FH (2007) Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach. *Molecular Ecology*, **16**, 3617–3633.
- Rogers RA (1986) Population differences in quantitative characters as opposed to gene frequencies. *The American Naturalist*, **127**, 729–730.
- Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genomic scans towards the characterization of candidate loci under parallel directional selection in these lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.
- Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae). *Molecular Biology and Evolution*, **24**, 1423–1438.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- Rueffler C, Van Dooren TJM, Leimar O, Abrams PA (2006) Disruptive selection and then what? *Trend in Ecology & Evolution*, **21**, 238–245.
- Rundle H, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
- Rundle HD, Schluter D (2004) Natural selection and ecological speciation in sticklebacks. In: *Adaptive Speciation* (eds Dieckmann U, Doebeli M, Metz JAJ, Tautz D), pp. 192–209. Cambridge University Press, Cambridge, UK.
- Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Molecular Ecology*, **12**, 315–330.
- Santiago E, Caballero A (1998) Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics*, **149**, 2105–2117.
- Savolainen V, Anstett MC, Lexer C *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441**, 210–213.
- Scheffer SJ, Hawthorne DJ (2007) Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera: Agromyzidae): apparent discordance among marker systems. *Molecular Ecology*, **16**, 2627–2637.
- Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753–763.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.
- Schluter D, Nagel LM (1995) Parallel speciation by natural selection. *American Naturalist*, **146**, 292–301.
- Scotti-Saintagne C, Mariette S, Porth I *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.] *Genetics*, **168**, 1615–1626.
- Shaw KL (2002) Conflict between mitochondrial and nuclear DNA phylogenies of a recent species radiation: what mitochondrial reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences, USA*, **99**, 16122–16127.
- Sheppard PM (1953) Polymorphism, linkage and the blood groups. *The American Naturalist*, **87**, 283–294.
- Sinervo B, Svensson E (2002) Correlational selection and the evolution of genomic architecture. *Heredity*, **89**, 329–338.
- Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**, 393–430.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Smith TB, Calsbeek R, Wayne RK, Holder KH, Pires D, Bardeleben C (2005) Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. *Journal of Evolutionary Biology*, **18**, 257–268.

- Steinfartz S, Weitere M, Tautz D (2007) Tracing the first step to speciation: ecological and genetic differentiation of a salamander population in a small forest. *Molecular Ecology*, **16**, 4550–4561.
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Storz JF, Nachman MW (2003) Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution*, **57**, 2628–2635.
- Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris* – large effective population sizes and rates of long-term gene flow. *Evolution*, **62**, 1936–1950.
- Svensson EI, Abbot JK, Gosden TP, Coreau A (2008) Female polymorphisms, sexual conflict and limits to speciation processes in animals. *Evolutionary Ecology*, in press.
- Sword GA, Joern A, Senior LB (2005) Host plant-associated genetic differentiation in the snakeweed grasshopper, *Hesperettix viridis* (Orthoptera: Acrididae). *Molecular Ecology*, **7**, 2197–2205.
- Thibert-Plant X, Hendry AP (2008) Five questions on ecological speciation addressed with individual based simulations. *Journal of Evolutionary Biology*, **22**, 109–123.
- Tsumura Y, Kado T, Takahashi T, Tani N, Ujino-Ihara T, Iwata H (2007) Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. *Genetics*, **176**, 2393–2403.
- Turner JRG (1967a) On supergenes. I. The evolution of supergenes. *The American Naturalist*, **101**, 195–221.
- Turner JRG (1967b) Why does the genotype not congeal? *Evolution*, **21**, 645–656.
- Turner TL, Hahn MW (2007) Locus- and population-specific selection and differentiation between incipient species of *Anopheles gambiae*. *Molecular Biology and Evolution*, **24**, 2132–2138.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLOS Biology*, **3**, 1572–1578.
- Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics*, **179**, 455–475.
- Vandepitte K, Jacquemyn H, Roldan-Ruiz I, Honnay O (2007) Landscape genetics of the self-compatible forest herb *Geum urbanum*: effects of habitat age, fragmentation and local environment. *Molecular Ecology*, **16**, 4171–4179.
- Vasemagi 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.
- Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.
- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.
- 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.
- Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.
- Vines TH, Schluter D (2006) Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 911–916.
- Vines TH, Köhler 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.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.
- Vitalis R, Dawson K, Boursot P, Belkhir K (2003) DetSel 1.0: a computer program to detect markers responding to selection. *Journal of Heredity*, **94**, 429–431.
- Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics*, **147**, 1091–1106.
- Whitlock MC (1999) Neutral additive genetic variance in a meta-population. *Genetical Research Cambridge*, **74**, 215–221.
- 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.
- 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.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Wright S (1940) Breeding structure of populations in relation to speciation. *The American Naturalist*, **74**, 232.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
- Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.
- Wu I, Ting C-T (2004) Genes and speciation. *Nature Review of Genetics*, **5**, 114–122.
- Xie X, Rull J, Michel A *et al.* (2007) Hawthorn-infesting populations of *Rhagoletis pomonella* in Mexico and speciation mode plurality. *Evolution*, **61**, 1091–1105.
- Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883–1893.
- Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences, USA*, **105**, 3421–3426.

The authors all use a combination of experimental and molecular genetic approaches to study speciation. Patrik Nosil and Daniel Funk's research investigates insect speciation in response to adaptation to different host-plant species. Patrik Nosil's work focuses on walking-stick insects in California while Daniel Funk's work focuses on leaf beetles in eastern North America. The two authors now collaborate on these systems. Daniel Ortiz-Barrientos' doctoral work treated the genetic basis of reinforced mating preferences in *Drosophila*. His current work involves speciation, ecological genetics, and sexual selection in flowering plants.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Conceptual models for the growth of genomic islands of divergence. In all cases, bars represent chromosomes, white boxes within them represent differentiated regions of the genome, and filled, black areas represent undifferentiated regions. Two-headed arrows represent regions of the genome where genetic exchange between populations is high. (A) Allopatric model. Divergence proceeds unimpeded by gene flow, with the proportion of the genome differentiated between two populations being positively related to time since divergence. (B) Ecological model. A new mutation (grey box within chromosome) arising near genomic regions under selection, and thus undergoing reduced introgression, has a higher likelihood of differentiating between populations than a new mutation arising in a region distant from those under selection. (C) Structural model. A new mutation (grey box within chromosome) arising near an inversion has a higher likelihood of differentiating between populations than a new mutation arising in a region distant from the inversion.

Table S1 Robustness of studies listed in Table 2 of the main text. 'Multiple methods' indicates whether more than one primary type of analysis was used to detect outliers (N=No; Y=Yes, one program/simulation method was run in different ways; Y+ = Yes, more than one program or statistical procedure was used). This column refers to the actual implementation of different programs, whereas other ways of confirming outlier status (e.g. replication of outlier status across different population pairs) are dealt with in subsequent columns. 'Quantiles' indicates the threshold of expected neutral differentiation used to determine whether a locus was an outlier. 'Replication across population pairs' indicates whether the replication of outlier status across multiple population pairs was evaluated (N = No; Y = Yes; Y+, direct = Yes, and the study also explicitly evaluated if outliers were associated with a specific ecological variable, for example by being outliers only in comparisons between population pairs that differ in that variable, and never outliers in population pairs similar for that variable; Y+, indirect = same as Y+, but although an association of outliers with an ecological variable was evident in the data, it was not explicitly noted). 'Type I error' indicates whether this type of error was

accounted for (N = No; Y = Yes, by correcting for the number of loci within a comparison (e.g. via Bonferroni correction); Y+ = Yes, via additional consideration of the number of population pairs in which the locus was an outlier). 'Mutation rate variation' refers to whether this possible confounding factor was discussed (N = No; Y = Yes, by arguing that gene flow negates the effects of mutation rate variation, or by comparing differentiation between regions known to differ in mutation rate; Y+ = Yes, by evaluating outliers that were associated with a specific ecological variable, a pattern unlikely to arise via mutation rate variation. We note here only studies that actually discussed this issue, but any study examining parallel divergence, and particularly those that noted associations between outliers and ecological variables, indirectly argue that mutation rate variation is an unlikely cause of outlier behaviour). Past studies suggest that divergence-based methods for detecting divergent selection are robust to demographic variability (Beaumont & Balding 1996), but we further note whether the potential confounding effects of 'demography' were discussed. 'Background selection' refers to whether it was considered. When the discussion of a particular factor was particularly explicit, we note the relevant page number.

Table S2 Summary of studies providing information on isolation by adaptation (IBA), where IBA refers to a positive association between the degree of adaptive phenotypic (or ecological) divergence and the level of genetic differentiation (here, at putatively neutral genetic markers). Provided is the study system, a description of the taxa or ecological forms examined, the type and number of markers used, the study design, and the main result with respect to IBA and IBD (where applicable). When reporting findings for IBA, the results reported are independent of geographic distance. Three main types of studies are reported: (i) population genetic studies explicitly examining IBA, generally using distance matrices of adaptive and neutral genetic divergence, (ii) population genetic studies using an AMOVA framework (Excoffier *et al.* 1992), where we report the per cent of total genetic variation observed between ecological types vs. that observed among populations within ecological types, and (iii) illustrative examples of a pattern consistent with IBA in a mosaic hybrid zone and in a phylogeographic study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.